8-5-2020

An Experimental Study on the Acquisition of Impersonals in Brazilian Portuguese

Karina Bertolino
University of Connecticut - Storrs, karina.bertolino@uconn.edu

An Experimental Study on the Acquisition of Impersonals in Brazilian Portuguese

Karina Bertolino, Ph.D.

University of Connecticut, 2020

This dissertation focuses on the acquisition of null impersonal structures in Brazilian Portuguese (BP), a partial null-subject language (PNSL). PNSLs allow definite null subjects under more restricted conditions than consistent null-subject languages (CNSLs). However, PNSLs are characterized by having impersonal constructions with generic null subjects which are absent in CNSLs.

Using the Truth Value Judgment Task, I investigated whether BP-speaking children know that they are acquiring a PNSL, in which null subjects should be understood as generic. The results show that children as young as 4-years-old correctly reject the definite reading of null subjects in impersonal structures and correctly accept their generic reading. The data suggest that BP-speaking children have early knowledge that they are acquiring a PNSL.

Using the Felicity Judgement Task, I tested children's knowledge of restrictions imposed on the well-formedness of impersonal structures in BP. Particularly, impersonal structures in BP are completely well-formed only when they have an overt marker of genericity (e.g., a deontic modal or an impersonal clitic *se*). 7-year-olds and 6-year-olds exhibited sensitivity to the adult grammar. 5- and 4-year-olds exhibited worse performance than the other groups, over-accepting the null subject in structures without an overt marker of genericity. It seems that at the age of 5-

years-old children did not completely acquire the impersonal pronoun *se*. This can be explained by the fact that impersonal *se* is not frequent in the input.

Building upon Holmberg's (2010a) Null Subject Parameters, I propose a learning model to address the problem of learnability that PNSLs and other (non-)null-subject languages impose. There is no evidence of missetting of the Null Subject Parameters by children acquiring any (non-)null-subject language, supporting the hypothesis that parameters are set in the earliest observable stages (Wexler 1995).

An Experimental Study on the Acquisition of Impersonals in Brazilian Portuguese

Karina Bertolino

BA., University of São Paulo, 2011

MA., University of São Paulo, 2013

MA., University of Connecticut, 2017

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2020

APPROVAL PAGE

Doctor of Philosophy Dissertation

An Experimental Study on the Acquisition of Impersonals in Brazilian Portuguese

Presented by

Karina Bertolino, B.A., M.A.

Major Advisor _____
Diane Lillo-Martin

Associate Advisor_____
William Snyder

Associate Advisor_____
Željko Bošković

University of Connecticut

2020

Acknowledgements

I wish to express my deepest gratitude to my advisor, Diane Lillo-Martin, without whom this dissertation would not be completed. I was fortunate to have as an advisor someone so knowledgeable and patient as her. Diane also had a lot to teach on the topic of this dissertation: acquisition of null subjects. My meetings with Diane played a crucial role in the completion of this dissertation, not only for their quality but also for the frequency in which she was available to discuss my ideas. I have told many of my colleagues about the importance of my weekly meetings with Diane, mainly when I was struggling with a low mood: I started writing what I could and as much as I could every day no matter how I felt, because I was going to meet with Diane and I wanted to have something to show her. I also would like to thank her for carefully reading all my drafts and correcting mistakes that I would otherwise have left in.

My other committee members have a lot to be thanked for as well. I enjoyed every discussion with William Snyder about the analysis of the data. Many of my first statistical analyses were revised after meeting with him. William was also very good at identifying flaws in reasoning and confusing passages in the dissertation which led to many revisions. I also had Željko Bošković as my associate advisor. I never thought someone could read a dissertation so quickly and still be able to comment on every detail. Željko helped me improve theoretical aspects of this dissertation. I incorporated his suggestions and ended up with a better version than I first had. I had some complicated and unnecessary passages in my first draft which he convinced me to get rid of. Without his help I would be telling a much more complicated story than I am telling here.

I also want to express my gratitude to Magdalena Kaufmann who read three chapters of this dissertation even though she was not part of my committee. She has a meticulous attention to detail and she is honest when improvement is needed, which is why I asked her to read my chapters.

I am also deeply thankful to Ian Roberts who was the first person to guide me through the study of null impersonal structures. Without him, this dissertation would be about something else.

I thank Letitia Naigles for teaching me how to conduct IPL experiments and analyze IPL data. Thank you for giving me this opportunity, Letty!

I would not have started doing research if it were not for Elaine Grolla, who was my advisor back in Brazil and taught me so much about language acquisition. I have known Elaine for many years and she gave me great support in my academic life. Needless to say, I am indebted to her. Elaine, eu tenho certeza que você vai ficar muito feliz quando eu te contar que terminei a minha jornada na UConn. Mal posso esperar!

I would like to thank all other students from the Department of Linguistics at UConn. Mainly Renato Lacerda for putting up with me everywhere. I could never thank him enough. Thank my dear classmates Emma Nguyen, Adrian Stegovec, Ryosuke Hattori, Jayeon Park and Roberto Petrosino. There are also other graduate students who I had the pleasure to have great conversations with: Pavel Koval (Pasha), Sabine Laszakovits, Margaret Chui Yi Lee, Shengyun Gu, Yuta Tatsumi and Marcin Dadan. I also want to thank Yoshiki Fujiwara, Kangzheng Gao, Shuyan Wang and Corina Goodwin for their feedback in the Acqui Lab presentations.

I should also thank Cynthia DeRoma who picked me up in New Haven when I got lost the first time I arrived in the US! Obrigada também por aquela conversa no dia em que almoçamos juntas, Cynthia.

I definitely need to thank Aida Talić who was extremely kind to let me sleep, eat and shower in her apartment for two weeks (!) when I first arrived in Connecticut.

Special thanks go to Tamara Cohen who took care of the administrative tasks and made my life so much easier.

This research would not have been possible without the children who participated in the study. Many thanks to the children from Museu da Imaginação, Grapali Brinquedoteca, Terra do Nunca Brinquedoteca and Play Space. Many thanks also to my niece Helena and children from other places who participated in the experiments.

I would like to thank Sonia Singh for her friendship and her help when I conducted the experiments in Brazil. Even without speaking Portuguese, she patiently accompanied me to some of the places in São Paulo where I conducted my study. Sonia helped me keep the data organized. Thanks to her I came back to the US without losing consent forms and answer sheets!

Many thanks to my housemate Joanna Glover. I will miss her a lot when I leave the US, as rare were the days in which she did not make me laugh. She is one of the smartest and funniest persons I have ever met and I treasure her friendship.

I also want to thank Laura. As I trusted her, I changed. Without this change, I would not have written this dissertation.

Table of Contents

Chapter 1: Introduction

This thesis focuses on the acquisition of null impersonal constructions with generic interpretation in Brazilian Portuguese (BP), a partial null-subject language. As we see in the sentence in (1), null impersonal structures have a non-realized subject which is usually interpreted as generic[1] (similar to the English pronoun 'one' or the generic 'you'):

(1) Em Lisboa *e* não pode beber mais chá. *BP*
In Lisbon not can drink:INF more tea
'In Lisbon one cannot drink tea anymore.'

According to Holmberg et al. (2009), the presence of generic null subjects is one of the characteristic properties of partial null-subject languages. Partial null-subject languages allow definite null subjects under more restricted circumstances than consistent null-subject languages (e.g., Italian, European Portuguese (EP), Arabic, Greek and Spanish). As will be discussed below, partial null-subject languages allow null subjects only in embedded clauses. Nevertheless, partial null-subject languages allow generic null subjects which are illicit in consistent null-subject languages. In consistent null subject languages, the null subject in (1) can only have the definite reading. In order to express the generic reading of the subject, consistent null-subject languages resort to overt strategies. For example, Romance languages resort to the generic *se*/*si* pronoun, as in the sentence in (2) from EP.

[1] In some contexts, impersonal constructions in BP also have the existential reading. See chapter 2 for discussion.

(2) Em Lisboa não *se* pode beber mais chá. *EP*

In Lisbon not SE can drink:INF more tea

'In Lisbon one cannot drink tea anymore.'

Consistent and partial null-subject languages are differentiated by the properties in (3). These generalizations, however, only apply to third person pronouns: partial null-subject languages might be more permissive regarding the use of null first and second person definite pronouns. As we shall see in Chapter 2, the mechanisms which derive third person definite pronouns are different from those which derive first and second person pronouns.

(3)a. *Consistent null-subject languages (CNSLs):*

Null definite subject pronoun (null 'he/she')

No null generic pronoun (null 'one').

Partial null-subject languages (PNSLs):

Null definite pronouns only if locally c-commanded by an antecedent.

Null generic subject pronoun.

(4) Em Lisboa *e* não bebe mais chá. *CNSLs* *EP*

In Lisbon not drink:3SG more tea

= 'In Lisbon s/he does not drink tea anymore.'

≠ 'In Lisbon one does not drink tea anymore.'

(5)a. Em Lisboa *e* não bebe mais chá. *PNSLs* *BP*

In Lisbon not drink:3SG more tea

= 'In Lisbon one does not drink tea anymore.'

≠ 'In Lisbon s/he does not drink tea anymore.'

(5)b. O João disse que *e* sabe inglês muito bem. *BP*

The John said:3SG that know:3SG English very well

'John said that he knows English very well.'

The null subject typology is more complex than was initially thought. The original formulation of the Null Subject Parameter (Rizzi 1982, 1986) only makes a distinction between consistent and non-null-subject languages. That is, this initial formulation implies that languages can either allow the subject of a sentence to be unexpressed as in Italian (6), or require it to be expressed, like English (7):

(6) (Tu) sei bella! *Italian*

you be:2SG beautiful

'You are beautiful.'

(7) *(You) are beautiful. *English*

Rizzi's Null Subject Parameter cannot account for the distribution of null subjects in the complex null subject typology presented in Table 1, nor can it provide a learning model for the acquisition of these languages.

Table 1: Summary of the (non-)null-subject typology

	Consistent null-subject language	**Partial null-subject language**	**Semi null-subject language**	**Non-null-subject language**
Example	*EP, Italian*	*BP, Finnish*	*Icelandic (I), Kriyol (K)*	*English, French*
Matrix subjects	Null	Overt	Overt	Overt
Embedded subjects	Null	Null	Overt	Overt
Generic subjects	Overt	Null	Null (I) Overt (K)	Overt
Non-thematic subjects	Null	Null	Null	Overt

* In most cases, null means "optionally null".

Adapted from Holmberg and Sheehan (2010: 132)

As shown in Table 1, the null subject typology includes consistent, partial, semi and non-null-subject languages. Consistent null subject languages allow every kind of null subject, except generic null subjects. Partial null-subject languages, allow every kind of null subject, except definite null subjects in matrix clauses. Semi null-subject languages only allow non-thematic subjects, but Icelandic has null generic subjects. Non-null-subject languages do not have any kind of null subject pronoun. Additionally to these languages, there are radical pro-drop languages, which allow every kind of null argument (e.g, Chinese, Korean, Thai).

In order to account for the null subject typology discussed above, Holmberg (2010a) proposes three parameters: (a) D in T, (b) P in T and (c) ϕ-dependent EPP parameter. The D in T parameter determines whether a languages allows null subjects in matrix clauses or not. In languages with a D(efiniteness) feature in T, third person null subject pronouns are interpreted as definite, as seen in the example in (8) from European Portuguese. In languages without D in T, third person null subject pronouns in matrix clause can only be interpreted as generic, as seen in the example in (9), from Brazilian Portuguese.

(8) Nessa escola *e* tem que fazer lição. *EP*
In.this school have:3SG that do:INF assignment
'In this school s/he has to complete assignments'

(9) Nessa escola *e* tem que fazer lição. *BP*
In.this school have:3SG that do:INF assignment
'In this school one has to complete assignments'

The parameter P in T accounts for the distribution of null subjects in embedded clauses and for presence of absence of non-thematic null subjects in a language. In languages with P in T, [Spec, TP] should be pronounced. Therefore, languages with P in T disallow null subjects in embedded clauses, as seen in the example in (10) from English. Languages without P in T can have [Spec, TP] pronounced or not, as seen in the example in (11) from BP.

(10) John said that *(he) knows Portuguese very well.

(11) O João disse que (ele) sabe inglês muito bem. *BP*

The John said:3SG that know:3SG English very well

'John said that he knows English very well.'

The ϕ-dependent EPP parameter defines whether subjects check the EPP feature in T or another element (e.g., adverbials) can fulfill this requirement. English is an example of language with a positive setting for the ϕ-dependent EPP parameter, since only DPs and pronouns assigned nominative case can appear in [Spec, TP]. Finnish, on the other hand, is a language with a negative value for the ϕ-dependent EPP parameter, since elements such as adverbials can appear in [Spec, TP], as seen in the example below.

(12) Tässä istuu mukavasti.

here sit:3SG comfortably

'One can sit comfortably here.'

The first research question investigated in this thesis concerns the structure of null impersonals in BP. According to Roberts (2010) and Holmberg (2005, 2010a,b), null pronouns are ϕP pronouns, that is, defective pronouns, which on their own cannot be definite. The generic interpretation of null pronouns in BP is explained by the fact that the language does not have a D-feature in T.

According to Holmberg (2010a), all languages which allow generic null subjects are ϕ-independent languages. That is so because the ϕP pronoun is the non-head of a chain formed by

this pronoun [3SG, NOM] and the features of T [T, 3SG, NOM]. This pronoun is deleted, as spelled out in (13), and as such it cannot check the EPP. If there is in fact an EPP feature that should be checked in [Spec, TP], it follows that sentences with generic null subjects need an element other the subject in [Spec, TP] to satisfy the EPP.

(13)a. [T, uφ, NOM] [vP [3SG, uCase] *v*...] → *Partial null-subject languages*

b. [T, 3SG, NOM] [vP [3SG, NOM] *v*...] →

c. [T, 3SG, NOM] [vP [~~3SG, NOM~~] *v*...]

However, in Chapter 2 and 3, I argue that no element is necessary to satisfy the EPP in impersonals in BP, as structures such as (14) are grammatical in this language:

(14) Não pode largar a escola. *BP*

Not can leave:INF the school

'One cannot drop out of school.'

In Chapter 2 and 3, I also investigate the hypothesis that BP requires an overt marker of genericity in impersonal constructions, such as deontic modals (15a) and the generic *se* pronoun (15b). As will be discussed in Chapter 3, this initial hypothesis has proved too strong, as sentences without overt markers of genericity (15c) are not completely ungrammatical for BP-speakers. However, BP-speakers have a *preference* for generic impersonal structures with an overt marker of genericity.

(15)a. Nessa escola *e* não **pode** trazer brinquedo.

In.this school not can:3SG bring:INF toy
'In this school one can bring toys.'

b. Nessa escola **se** traz brinquedo.
In.this school SE bring:INF toy
'In this school one brings toy'

c. ?Nessa escola *e* traz brinquedo. *BP*
In.this school bring:3SG toy
'In this school one brings toys.'

In Chapter 4, I investigate whether children acquiring BP know that null subjects in impersonal constructions have the generic reading rather than the definite one. In other words, this chapter addresses the question of whether BP-speaking children know that they are acquiring a partial null-subject language instead of a consistent null-subject language. In order to answer this question I conducted an experiment using a Truth-Value Judgement Task (Crain and McKee 1985). The results show that Brazilian children as young as 4-years-old know they are acquiring a partial null-subject language.

Chapter 5 discusses an experiment using a Felicity Judgement Task which was conducted to investigate the contexts in which children allow null impersonals in BP. In this task, a pair of sentences with a generic null subject and with a *se* pronoun were presented to children and they had to judge which sentence in the pair sounded better. The main question that the experiment aimed to investigate was whether children prefer impersonal sentences with an overt marker of genericity as adults do. It was found that 4- and 5-year-olds exhibited worse performance than 6- and 7-year-olds. Particularly, young children had problems choosing the sentence with *se* when it

was the expected answer. As impersonal *se* is not the preferred strategy of impersonalization in BP, it is presumably not frequent in the input children receive. This can explain the late acquisition of impersonal *se*.

In Chapter 6, I propose a learning model for children acquiring partial null-subject languages and other (non-)null-subject languages. This learning model builds upon Holmberg's (2010a) (a) D in T, (b) P in T and (c) ϕ-dependent EPP parameters. I deduce the initial value of these parameters from the Subset Principle (Berwick 1982; Wexler and Manzini 1987). After deducing the initial values, I discuss which sort of evidence in the input can be used by the child to switch the initial value of these parameters when needed in different (non-)null subject languages.

To sum up, this thesis aims to answer the following questions:

I. What is the syntactic structure of null impersonals with generic reading in BP? Which constraints are imposed on the well-formedness of these sentences? *(Chapter 2)*

II. Based on empirical evidence, can we conclude that impersonal structures in BP (i) have a EPP feature which needs to be checked by an element other than the null generic pronoun? Can we also conclude, based on empirical evidence, that (ii) the generic reading of impersonal sentences requires an overt marker of genericity? *(Chapter 3)*

III. Do children acquiring BP know that null subjects in impersonal constructions have the generic reading rather than the definite one? That is to say, do BP-speaking children know they are acquiring a partial null-subject languages instead of a consistent null-subject language? *(Chapter 4)*

IV. In which contexts do children allow null impersonals in BP? Do they obey the same restrictions imposed by the adult grammar? *(Chapter 5)*

V. How do children acquire different (non-)null-subject grammars? *(Chapter 6*)

Chapter 2: Impersonal Structures in BP

0 Introduction

The main objective of this chapter is describe the structure of impersonal sentences in Brazilian Portuguese (BP), especially null impersonals with a generic reading.

In section 1, I first examine the nature of the null subjects more generally: I compare three hypotheses: (i) null subject as *pro*; (ii) the pronominal agreement hypothesis; and (iii) the deletion hypothesis. I adopt the third hypothesis which claims that null subjects are pronouns deleted at PF.

Particularly, I assume, following Roberts (2010) and Holmberg (2010a), that consistent null-subject languages have an uninterpretable D-feature on T. This D-feature is responsible for the definite interpretation of null subjects. Partial null-subject languages, such as BP, do not have a D-feature on T, and hence definite null subjects are generally disallowed. Partial null-subject languages allow generic null subjects in impersonal sentences because they lack a D-feature on T.

In section 2, I discuss the historical changes that BP underwent towards a non-pro-drop parameter setting and the status of definite null subjects in the language. In section 3, I explain the different types of non-definite subjects and specify which ones are important for the impersonal structures that I am investigating here. In section 4, I discuss whether impersonal structures in BP need to have [Spec, TP] filled by an adverbial in order to be well-formed. In section 5, I discuss the role of deontic modals and the clitic *se* as overt expressions of genericity. Section 6 concludes the chapter.

1 The Null Subject

If one assumes that all theta roles must be assigned, it follows that in the structure below, some null syntactic category should receive the external theta-role that *correram* ('ran') needs to assign in (1):

(1) Correram por dez quilômetros. *European Portuguese*

Ran:3PL for ten kilometers

'They ran ten kilometers.'

Another theoretical argument supporting the existence of a null syntactic category comes from the EPP. The EPP, to be discussed in more detail below, imposes a requirement to have the subject position [Spec, TP] filled in all sentence types (Chomsky 1981, 1982)[2]. If the EPP holds, it follows that there is a null element in [Spec, TP] in sentences like (1) and (2), since there is no other element available to fulfill this requirement.

(2) Sono andati tutti via. *Italian*

Are gone all away

'They have all gone away.'

(Roberts 2010: 72)

[2] Throughout this thesis, I use the term EPP pretheoretically, without assuming that the EPP requirement exists. Considering the alternative that the EPP does not exist, its effects could be derived from independent conditions of the grammar (Bošković 2002).

Non-pro-drop languages, such as English, provide a more straightforward argument for the EPP. As non-pro-drop languages have no null-subject, the sentence in (3a) is ungrammatical. The sentence in (3a) becomes grammatical if an expletive is added (3b), filling [Spec, TP] and satisfying the EPP.

(3)a. *Is likely someone will leave.

(3)b. It is likely someone will leave.

The exact nature of the empty category in (1) and (2) has been subject to debate, and three basic answers have been proposed (Camacho 2013). One largely accepted answer is that the empty category in tensed clauses in consistent null-subject languages (e.g., Italian, Spanish and European Portuguese) is *pro* (Chomsky 1981; Rizzi 1982, 1986; Rizzi and Shlonsky 2007, among others). Another proposal assumes that agreement on the verb can satisfy the EPP in virtue of being pronominal, eliminating the need for a separate syntactic null subject element (Jelinek 1984; Borer 1986; Ordóñez 1997; Alexiadou and Anagnostopoulou 1998; Kato 1999; Ordóñez and Treviño 1999; Barbosa et al. 2005; Barbosa 2010; Sigurðsson 2011, among others). Roberts (2010) and Holmberg (2005, 2010a) propose that the definite null subject in (1) and (2) are deleted φPs. In the following sections I will consider these proposals, and how they apply to BP.

1.1 The Null Subject as pro

Building on earlier work by Chomsky (1981; 1982), Rizzi (1986) proposes that the empty category in sentences like (1) and (2) is a silent pronoun called *pro*. *Pro* has all the features of an overt DP subject, such as nominative case and number.

In Rizzi's proposal *pro* is subject to two requirements: (i) it needs to be *licensed* under head-government and (ii) the content of *pro* needs to be *recovered* by rich agreement:

(4) *Pro-drop parameter*

(i) *pro* is governed by X^0_y;

(ii) Let X be the licensing head of an occurrence of *pro*: then *pro* has the grammatical specification of the features on X coindexed with it.

(Rizzi 1986: 519-20)

The pro-drop parameter states that *pro* should be governed by a head X^0 of type y. Whether a language has this head is a language specific property. The choice of X^0 is subject to cross-linguistic variation. In languages such as Italian, European Portuguese (EP) and Spanish, INFL is the relevant choice for X^0 and the value for y includes the inflection. In non-null-subject languages such as English, no value would be assigned to y ($\{X_y\}$ is empty), and therefore there is no category that licenses *pro*.

In the following sentence from EP, *pro* would be licensed because it is governed and assigned Case by I(NFL), which would qualify as X^0_y in the language. *Pro* would be inherently unspecified for φ-feature values and it would inherit the grammatical specifications of I(NFL), with which it would be coindexed:

(5)a. Foram embora. *EP*

go:PAST:3PL away.

'They went away.'

b. [$_{IP}$ pro$_{i\ [3PL]}$ [$_{I}$ I$_{i\ [3PL]}$][$_{VP}$ foram embora]]

In Government and Binding Theory (Chomsky 1982), *pro* is a non-overt counterpart of pronouns. The equal specification of *pro* for the features [± anaphor] [± pronominal] compared with overt pronouns makes *pro* not a primitive concept, but an entity that can be deduced from more general aspects of the theory. It was observed that anaphors and NP-traces are subject to Principle A of Binding Theory, wh-traces and R-expressions are subject to Principle C and pronouns and *pro* are subject to Principle B. PRO, characterized by the features [+anaphor, +pronominal], would have no overt counterpart, since overt subjects need to be governed and PRO must be ungoverned.

Table 1: Typology of NPs

Type	**Overt**	**Non-overt**
[+anaphor, -pronominal]	Anaphor	NP-trace
[-anaphor, +pronominal]	Pronouns	pro
[-anaphor, -pronominal]	R-expressions	Wh-trace
[+anaphor, +pronominal]	-	PRO

However, within the Minimalism Framework, the existence of elements such as *pro* raises some issues. Holmberg (2005: 536) points out that *pro* cannot be maintained in a theory that makes a distinction between interpretable and uninterpretable features (Chomsky 1995). Φ-features can be interpretable or uninterpretable. The φ-features of NPs (or DPs) are interpretable since they restrict the denotation of the NP, while the φ-features of a predicate are uninterpretable, since they do not restrict its denotation. Consider the sentence in (6a), with the derivation in (6b):

(6)a. As trabalhadoras parecem ter sido contratadas.
The:FEM.PL worker:FEM.PL seem:3PL have been hired:FEM.PL
'The female workers seem to have been hired.'

b. [$_{TP}$ [as trabalhadoras]i [$_{T'}$ -m $_{[3.PL]}$ [$_{VP}$ parece- [$_{TP}$ t_i [$_{T'}$ ter [$_{VP}$ sido [$_{TP}$ t_i -das$_{[FEM.PL]}$
[$_{VP}$ contrata-t_i]]]]]]]]

In (6), the interpretable φ-features (gender and number) of the noun *as trabalhadoras* (the workers) enter into a checking relation with the uninterpretable φ-features of T. The uninterpretable features are eliminated in the course of the derivation in LF, although they might be visible in PF.

The important point for us is that given this theory, *pro* could not have its features specified (i.e., valued)) by T, since T is essentially unspecified in the relevant respect.

Pro is not only problematic from a theoretical point of view. Some languages that allow null subjects also pose an empirical problem for Rizzi's original proposal. Partial null-subject languages such as Modern Hebrew pose a problem in that I(NFL) can be the X that will license a null subject, but in a way that is restricted. In Modern Hebrew, null subjects in matrix clauses are allowed only in the future and past tense and they are further restricted to first and second person (Borer 1986: 392):

(7)a. Hu 'axal 'et ha-tapu'ax. *Modern Hebrew*

He eat:PAST.3SG ACC the-apple

'He ate the apple'

b. *'Axal 'et ha-tapu'ax.

Eat:PAST.3SG ACC the-apple

'He ate the apple.'

c. 'Axalti 'et ha-tapu'ax.

Ate:1SG ACC the-apple

'I ate the apple.'

d. 'Ani/'ata/hu 'oxel 'et ha-tapu'ax.

I/you/he eat:SG ACC the-apple

'I/you/he eat the apple.'

e. *'Oxel 'et ha-tapu'ax.

Eat:SG ACC the-apple

'I/you/he eat the apple.'

It is also unclear how the content of the null subject is recovered in *radical pro-drop languages* such as Japanese, Korean and Chinese in Rizzi's proposal. For Chinese, Huang (1984) observes that the language allows null subjects, although it entirely lacks AGR. The same holds for Korean and Japanese.

In sum, if we want to follow certain minimalist assumptions about the way agreement works, it seems to be problematic to consider *pro* a category present in the lexicon. Rizzi's classical approach to how *pro* is licensed and how its content is recovered faces empirical problems when we try to account for the data in partial null-subject languages and radical pro-drop languages.

1.2 Pronominal Agreement Hypothesis

The hypothesis that I(NFL) can be pronominal had its first formulations in Rizzi (1982: 143), Hale (1983), Jelinek (1984) and Borer (1986). Camacho (2013: 76) condenses Hale (1983)

and Jelinek's (1984) approach as in (8)[3]. The Pronominal Agreement Hypothesis is also known as 'I-subject' view, since agreement is located in I(NFL):

(8) Pronominal Agreement Hypothesis[4]

(i) AGR/INFL may license the EPP.

(ii) Morphological affixes can receive theta-roles.

Thus, a sentence with a null-subject like (9) would have the following derivation:

(9)a. Despert-aron. *Spanish*

awake:PAST.3PL

'They woke up.'

b. [$_{VP}$ despert- aron $_{\theta\ +D}$]

c. [$_{TP}$ [T + despertaron$_{+D,\ \theta}$ [$_{VP}$ t_v]]]

(Camacho, 2013: 77)

[4] Hale (1983) formulates his proposal based on Papago and Jelinek (1984) extends this proposal to Spanish.

In (9b) the pronominal person (third) and number (plural) morpheme -aron is assigned a theta-role by the verb (despert-). Notice that the D-feature indicates that the verb's inflectional ending is pronominal. Later the verb moves to T to check the EPP. If INFL is pronominal, it can satisfy both the Theta Criterion and the EPP.

Another version of this hypothesis is developed by Alexiadou and Anagnostopoulou (1998). According to the authors, there are two ways to satisfy the EPP and languages vary in the way it can be satisfied. That is, in Alexiadou and Anagnostopoulou's proposal, the EPP is parametric. In Germanic languages such as Icelandic and English, the EPP is satisfied by inserting an overt expletive (10) or by merging a DP (11). In Romance languages that happen to have rich verbal morphology such as Italian, European Portuguese or Spanish, the verb checks the EPP by moving to T (12).

(10)a. það lasu einhverjir stúdentar bókina. *Icelandic*

There read some students the.book

'There read some students the book.'

b. [[$_{IP}$ það$_{EPP}$ [$_{I}$ [$_{VP}$ lasu einhverjir stúdentar bókina]]]].

(11)a. Einhverji stúdentar lasu bókina.

Some students read the.book

'Some students read the book.'

b. [[$_{TP}$ Einhverji stúdentar$_{EPP}$ [$_{T}$ [$_{VP}$ lasu bókina]]]].

(12)a. Leyo Juan el libro. *Spanish*

Read Juan the book

'Juan read the book.'

b. [$_{TP}$ [$_{T'}$ T + leyo $_{+D}$ [$_{T}$ [$_{VP}$ Juan t_v el libro]]]].

(adapted from Alexiadou and Anagnostopoulou 1998: 492)

As we can see in the structure in (12b), languages like Spanish in which the verb's inflectional ending is pronominal have a D-feature on T that needs to be checked. The EPP requirement arises from the need to check the D-feature. Whenever the language lacks the D-feature on T, the EPP must be checked by a full expletive, as in (10) or a full XP, as in (11). For Alexiadou and Anagnostopoulou (1998), the locus of the EPP feature is AgrSP.

If agreement is pronominal, we expect to find definite properties associated with agreement. The following examples in Spanish suggest that this is true:

(13)a. Las mujeres tenemos esperanza. *Spanish*

The women have:1PL hope

'We women have hope.'

b. Las mujeres teneis esperanza.

The women have:2PL hope

'You women have hope.'

c. Las mujeres tienen esperanza.

The women have:3PL hope

'Women have hope.'

(Jelinek 1984: 48)

The examples in (13) suggest that definiteness is determined by the agreement marker and not by the subject. In (13a) through (13c), the subject is kept the same (*las mujeres*), but by the translations in English we see that the reference of the subject changes according to the verbal inflection. In (13a), the speaker is included in the set of women and this is expected if we consider that the first person plural inflection on the verb is the one that marks the reference. In (13b), the addressee is included in the set of women and the second person is marked as a verbal suffix, but not in the subject *las mujeres*. In (13c), the interpretation excludes the speaker and it can or not include the addressee, which is consistent with the third person plural marker on the verb.

However, as Camacho (2013: 79-80) points out, it is not always the case that agreement is definite. First of all, the paradigm above is not possible in Italian. Even in Spanish, the pattern seen in (13) is only possible with third person plural DPs. In (14), where the DP and the

inflection are singular, it is not possible to combine a third person singular DP and a first person singular inflection ((14b) is ungrammatical, contrasting with (13a)):

(14)a. El estudiante tiene mala memoria. *Spanish*

The student have:3SG bad memory

'The student has bad memory.'

b. *El estudiante tengo mala memoria.

The student have:1SG bad memory

'I the student have bad memory.'[5]

(Camacho 2013: 80)

Pronouns in Spanish also don't show the pattern in (13) described by Jelinek (1984), nor do pronoun + DP combinations, as we can see below in the examples in (15). We see in these examples that the verbal inflection cannot determine definitely when the subject is a pronoun (15c) or a pronoun + DP (15a). It is rather the subject that is determining the reference ((15b) and (15d)). These data are also problematic for the pronominal agreement hypothesis:

(15)a. *Nosotros los estudiantes tienen mala memoria.

We the students have:3PL bad memory

[5] In Camacho, the sentence in (14b) is translated as 'the students have bad memory'. I believe, though, that this translation is a typo, since the interpretation should include the speaker if the intention is to show that the inflection cannot be definite in this sentence. Also, there is no reason to translate *el estudante* as 'the students', as both the subject and the inflection are in the singular form in (14b).

'We the students have bad memory.'

b. Nosotros los estudiantes tenemos mala memoria.
We the students have:1PL bad memory
'We the students have bad memory.'

c. *Ellos (los estudiantes) tenemos mala memoria.
They (the students) have:1PL bad memory
'We the students have bad memory.'

d. Ellos (los estudiantes) tienen mala memoria.
They (the students) have:3PL bad memory
'They the students have bad memory.'

(Camacho 2013: 80)

Aside from the data in (14) and (15) that are problematic for the pronominal agreement hypothesis, there are other cases that challenge this account. One of them comes from Spanish data on ellipsis found in Saab (2009, 2010, 2012), as Camacho (2013) points out. The sentences in (16) show that since tense is an interpretable feature, the verb's tense under ellipsis must be identical to its overt counterpart.

(16)a. Los estudiantes aprendieron mucho chino y Antonio también ~~aprendieron mucho~~

The students learn:PAST much Chinese and Antonio also learn:PAST much ~~chino~~.

Chinese

'The students have learned a lot of Chinese and Antonio too.'

b. ??Los estudiantes aprendieron mucho chino en el pasado y Antonio en el futuro

The students learn:PAST much Chinese in the past and Antonio in the future

también ~~aprenderá mucho chino~~.

also learn:FUT much Chinese

'The students learned much Chinese in the past and Antonio in the future also will learn much Chinese.'

The same is not required for gender, since gender is an uninterpretable feature on the verb. As we can see in (17), the elided verb has the feminine marker, while the overt one has the masculine marker.

(17) Juan fue localizado en el restaurante y Marta también ~~fue localizada en el~~

Juan was located:MAS in the restaurant and Marta too was located:FEM in the ~~restaurante~~.

restaurant

'Juan was located at the restaurant and Marta too.'

(Camacho 2013: 82)

If agreement can be pronominal, person and number should be interpretable in the same way as tense. Therefore, the prediction is that a verb's person and number features elided under ellipsis should be identical to the person and number features on the overt verb. As can be seen in (18) the elided verb can actually have different features than the overt verb. Therefore the prediction is not corroborated by the data, suggesting that *agreement is not pronominal*:

(18) Juan fue al cine y nosotros también ~~fuimos al cine~~.
Juan went:3SG to.the movies and we also went:1PL the movies
'Juan went to the movies and we did too.'

(Camacho 2013: 82)

There are other cases that are problematic for the pronominal agreement hypothesis. For a complete presentation of these cases, see Camacho (2013: 78-86).

1.3 A New Typology of Null Subjects

In this section I turn to the analysis proposed by Roberts (2010) and Holmberg (2010a) of null subjects. They assume that *pro* is not an independent category, since that would not be compatible with minimalist assumptions. Instead, they propose that null subjects are φPs, that is, defective pronouns that on their own cannot be definite. What determines whether a null subject is definite or not is the presence of uD in T. In languages with uD in T, a φP pronoun can be

interpreted as definite. In languages without uD in T, such a null subject can only be interpreted as impersonal. Consistent null-subject languages, such as Italian, Spanish and European Portuguese (EP), have uD in T. Partial null-subject languages, such as Finnish and BP, do not have uD in T and the same is true for non-null-subject languages, such as English and French.

According to Holmberg (2010a), in *consistent null-subject languages*, an Aboutness-shift topic values the uD-feature of T. Holmberg (2010a) observes that third person null subjects in consistent null subject languages are dependent of an antecedent, which does not need to be overt. This antecedent is a particular type of topic, an Aboutness-shift topic (Frascarelli 2007). As shown in the sentence in (19a) from Italian, the pronoun cannot be null if there is a mismatch between the topic of the preceding sentence (i.e., the exhibition) and the reference of the null pronoun (i.e., Gianni) . In (19b), the pronoun can be null because it corresponds to the topic of the preceding sentence (i.e., Gianni).

(19)a. Questa mattina, la mostra è stata visitata di Gianni. Più tarde **e*/egli/lui ha
this morning the exhibition was visited by Gianni. Later he/he
visitato l'università.
visited the university.

b. Questa mattina, Gianni ha visitato la mostra. Più tarde *e*/egli/lui ha visitato
this morning Gianni visited the exhibition. Later he/he visited
l'università.
the university.

(Holmberg 2010a: 96)

The Aboutness-shift topic is base-generated in the C-domain, as represented in (20). This topic may or not be null.

(20) [$_{CP}$ <Gianni$_1$> [questa mattina Gianni$_1$ ha visitato la mostra]].

[$_{CP}$ <$\varnothing_2$> [questa mattina φP$_2$ ha visitato la mostra]].

1 = 2

(Holmberg 2010a: 96)

In (21), I spell out how third person definite null subjects are generated in *consistent null -subject languages*. As represented in (21a), finite T has unvalued φ-features. For this reason, T probes for a category with matching valued features. This category with matching valued features is the defective subject pronoun (i.e., φP) in [Spec, vP]. The φ-feature values of the pronoun are copied by T (i.e., which Roberts (2010) and Holmberg (2010a) refer as T incorporation): this is shown in step (21b). As seen in (21b), T also values the subject's unvalued case feature. Notice that now T shares all values of φP. That is to say, the T's feature values are a superset of φP's values. The probe and the goal in (21b) form a chain and as such they are subject to chain reduction. The principles of chain reduction are the following: (a) only one chain copy should be pronounced and (b) the highest chain copy is the one that is pronounced. Therefore, the subject φP is not pronounced, as represented by (21c). As the chain includes a D-feature which is valued by an Aboutness-shift topic, the result is a null definite pronoun.

(21)a. [T, D, uφ, NOM] [*v*P [3SG, uCase] *v*...] →

b. [T, D, 3SG, NOM] [*v*P [3SG, NOM] *v*...] →

c. [T, D, 3SG, NOM] [*v*P [~~3SG, NOM~~] *v*...]

According to Holmberg (2010a), the EPP in a structure like (21) is checked by the null Aboutness-shift topic in the C domain, which values [uD] in T, in the manner outlined in (22):

(22) Ha comprato una macchina nuova. *Italian*

has bought a car new

[$_{CP}$ <DP_1> [$_{TP}$ ha+$T_{[D1, 3SG, \sout{EPP}]}$ [$_{vP}$ <$\varphi P_{[3SG, NOM]}$> comprato…]]]

For first and second person null subjects in consistent null-subject languages, Holmberg (2010a) adopts Sigurðsson's (2004) and Frascarelli (2007) proposal that there are features representing the speaker and the addressee in every clause in the C-domain. In this manner, the speaker and the addressee are available as local antecedents for the null pronoun.

In *partial null-subject languages*, such as BP and Finnish, T does not have a uD-feature. In (23), we see the derivation of third person null subjects in *partial null-subject languages*. Finite T has unvalued φ-features in partial null-subject languages. T probes for a category with matching valued features, just like in consistent null-subject languages. T copies the φ-features values of the subject and, in return, the subject has its Case-feature valued (23b). T and the φP pronoun will form a chain and the subject will be deleted by chain reduction (23c). However, as partial null-subject languages do not have D in T, the interpretation of the subject can't be definite through this derivation, but only generic.

(23)a. [T, uφ, NOM] [*v*P [3SG, uCase] *v*...] →

b. [T, 3SG, NOM] [*v*P [3SG, NOM] *v*...] →

c. [T, 3SG, NOM] [*v*P [~~3SG, NOM~~] *v*...]

According to Holmberg (2010a), as generic null subjects do not have an Aboutness-shift topic as an antecedent, the EPP must be checked by another category in the sentence. According to Holmberg, the generic null subject cannot check the EPP because it is in [Spec, vP]. In partial null-subject languages, the EPP will be satisfied by categories such as adverbials, as shown in (24) and (25), for Finnish and BP, respectively. This is discussed in more detail in section 4.

(24) Tääla ei saa polttaa. *Finnish*

Here not may smoke

'One can't smoke here.'

(Holmberg 2005: 540)

(25) Aqui não pode fumar. *BP*

Here not can smoke.

'One cannot smoke here.'

Partial null-subject languages have definite null subjects when these are controlled by a subject in a higher finite clause, as we can see in (26) and (27) from Finnish and BP, respectively. Therefore, according to Holmberg (2010a), while definite null subjects in consistent null-subject

languages are incorporated φPs, definite null subjects in partial null-subject languages are DPs which are controlled by a higher subject in the matrix clause.

(26) Jari$_i$ sano että e_i istuu mukavasti tässä. *Finnish*
Jari says that sits comfortably here.
'Jari says that he sits comfortably here.'

(Holmberg 2010a: 102)

(27) João$_i$ me contou que e_i vende cachorro quente na praia. *BP*
João me told that sells dog hot at.the beach
'John told me that he sells hot dog at the beach.'

(Rodrigues 2004: 142)

Rodrigues (2004), Ferreira (2004) and Modesto (2000a) argue that the null subject in (26) and (27) is an instance of PRO. There is an independent issue of how traditional PRO should be analyzed; thus Hornstein (1999) treats it in terms of movement into a theta position. Holmberg and Sheehan (2010) argue that the control relation displayed by the subject in the matrix clause and the embedded null subject in partial null-subject languages is distinct from both obligatory control and non-obligatory control in non-finite clauses. Focusing on the differences between obligatory control and control in finite clauses in partial null-subject languages, Holmberg and Sheehan (2010) observe that the grammaticality of embedded null subjects is not regulated by Landau's (2004) control calculus (i.e., embedded null subjects are not restricted to clauses with

dependent tense), but by the requirement that they be bound by the closest c-commanding antecedent. The authors conclude that embedded null subjects in finite clauses are not PRO.

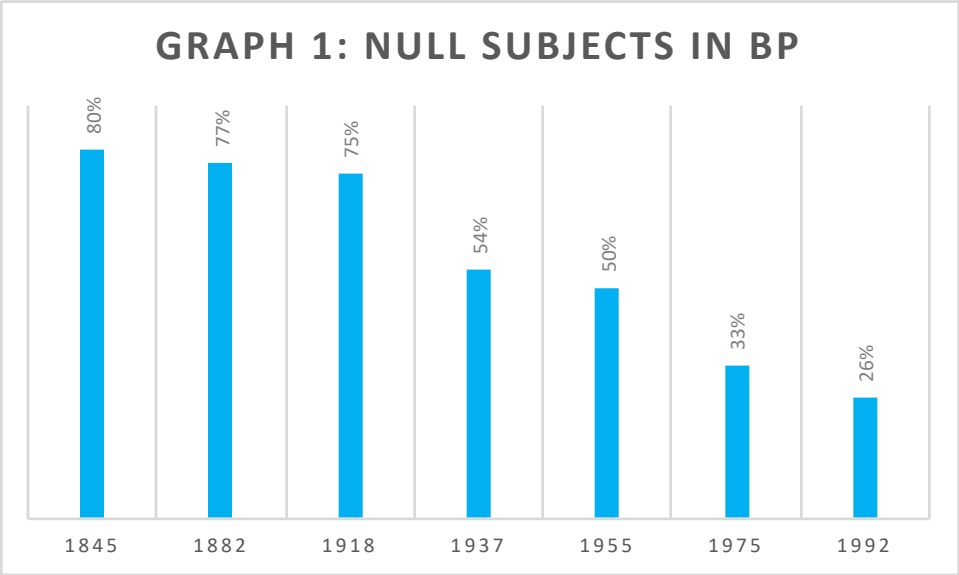
In the next sections, I discuss the behavior of null subjects in BP in detail, starting with the historical changes that affected BP towards a non-pro-drop status.

2 The Null Subject in BP

2.1 Historical Changes

In order to understand the current status of null subjects in BP, it is helpful to look at the changes in the language regarding the use of these subjects from the 19th century until recent years. Duarte (1993) conducted a diachronic research looking at plays written from 1845 to 1992. She presents a graph showing the occurrence of definite null subjects in BP throughout seven periods in history:

Graph 1: Proportion of definite null subjects in BP throughout the history



(adapted from Duarte 1995: 19)

As shown by Graph 1, BP has been gradually losing the null subject phenomenon. As Duarte (1993) argues, this gradual loss is strongly correlated with the change in the pronominal system of the language: the second person pronouns 'tu' (singular) and 'vós' (plural) were replaced by 'você' and 'vocês', respectively, which are treated as third person agreeing form; also, 'nós', first person plural, has been replaced by 'a gente', also a third person agreeing form. Changes in the verbal paradigm also happened, with the reduction of the feature *person*. Nowadays, the verbal paradigm of colloquial BP is composed only of three distinct forms, most of them being the unspecified form corresponding originally to the third person. As we can see in Table 2, by the 3rd period, the only grammatical person that does not present syncretism with other grammatical persons is the first person. Because of that, it could be assumed that the first person is the only one that carries the feature *person*. Nevertheless, in the past, the inflectional system of BP was richer, composed of six distinct forms, as a comparison between the columns named '1st period' and '3rd period' show. The 2nd period represents an intermediate stage that could still be found in the speech of older adults when Duarte (1995) conducted the study that we will describe in subsection 2.2.

Table 2: Changes suffered in the pronominal and inflectional paradigm in BP (*amar 'to love'*, present tense)

Pronouns	**1st Period** **(18th century)**	**2nd Period** **(first half of the 20th century)**	**3rd period** **(late 20th century on)**
Eu (1st person)	am**o**	am**o**	am**o**
Tu (2nd person)	am**as**	-	-
Você (2nd person)	ama	ama	ama

Ele/Ela (3rd person)	ama	ama	ama
Nós (1st person pl.)	ama**mos**	ama**mos**	-
A gente (1st person pl.)	-	ama	ama
Vós (2nd person pl.)	ama**is**	-	-
Vocês (2nd person pl.)	ama**m**	ama**m**	ama**m**
Eles/elas (3rd person pl.)	ama**m**	ama**m**	ama**m**

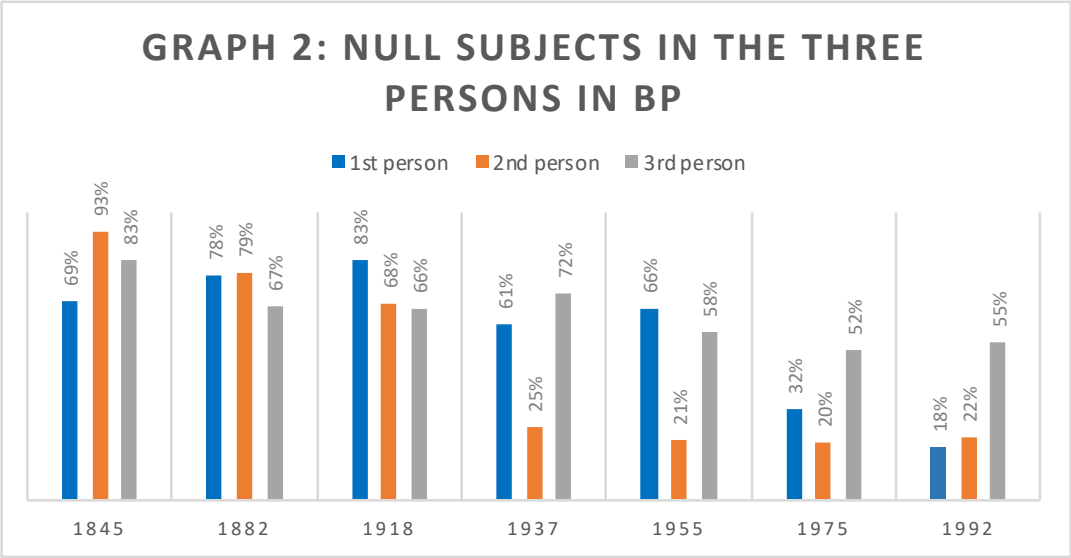
The table above shows that the first person singular is the only form that has a distinct inflection (**-o**). However, even the first person, which has preserved the feature 'person' in most dialects, may be losing its specification in a variety of BP. There is at least one variety of BP that does not mark the feature *person* at all. Lucchesi et al. (2009) notice that in a dialect of BP spoken in Bahia by Afro-descendants, there is the possibility to leave the first person singular unspecified, that is, the verb with the first person pronoun receives the agreement marker corresponding originally to the third person (-∅):

(28)a. Eu trabalha na roça.
I work:3SG in.the farm
'I work on the farm.'

b. Eu trabalhou na roça.
I work:3SG.PAST in.the farm
'I worked on the farm.'

Now, returning to definite null subjects, Graph 2 shows the percentage of null subjects in BP for each grammatical person throughout history. Percentages refer to the interpretation of the null subject (i.e., first, second, third person), rather than the verbal forms:

Graph 2: Occurrence of definite null subjects in the three persons in BP throughout the history



(adapted from Duarte 1995: 20)

The historical change more drastically affected the second person, with the percentage of null subjects being reduced from 93% in 1845 to 22% in 1992. We observe that in 1937 the percentage of null subjects drops to 25% in the second person. This coincides with the loss of the inflection **-s** used with the second person pronoun 'tu' and also with the loss of this pronoun in some dialects of BP. The pronoun 'tu' with the unspecified verb is still used in several regional dialects from Brazil, though. The option to express the second person with the pronoun 'você' became available in the 2nd period, as we can see in Table 2. Originally, 'você' was an honorific pronoun that triggered third person agreement and later it became grammaticalized as a second

person pronoun (Tarallo 1983). 'Você' as a pronoun also accompanies a verb that is unspecified for person.

The second person plural was also affected: 'vocês' replaced 'vós' completely and the verb started to exhibit a syncretism with the third person plural. Such a drastic change in the percentage of null subjects more strongly seen in the second person in Graph 2 can be explained by the fact that this is the person that was more affected by the changes in the pronominal system and in the verbal paradigm.

Duarte (1995) conjectures that the decreasing percentage of null subjects in the first person is correlated with the occurrence of 'a gente' to express first person plural. The pronoun 'a gente', as 'você', appears with an unspecified verbal form (see Table 2).

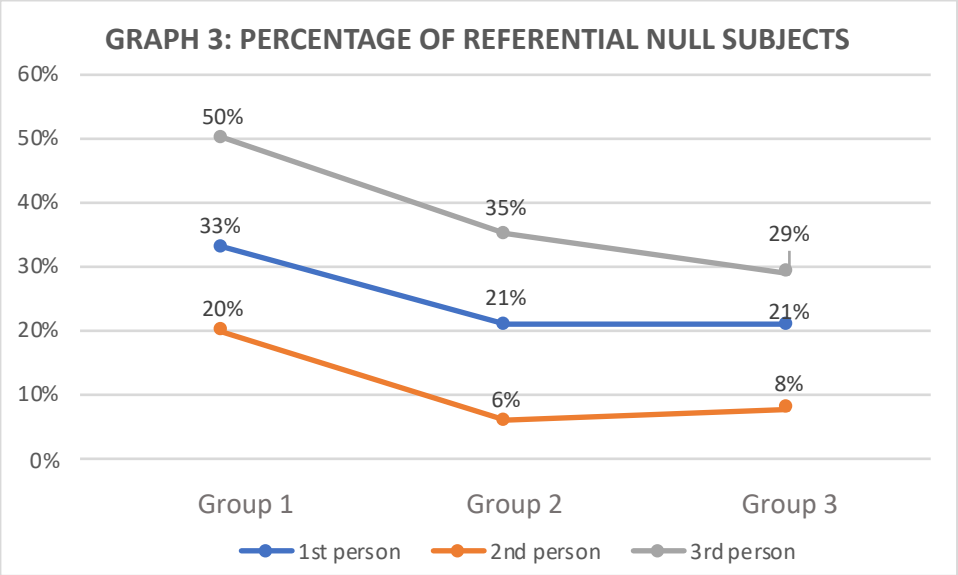
The third person was kept relatively stable. Looking at Table 2 we see that the pronouns and the inflections for the third person did not change over time. Once more, a change, or in the case of the third person, a lack of change in the verbal paradigm, would explain the pattern of null subjects found in BP. The small decrease in the percentage of null subjects observed in the third person could just been seen as an adjustment to a system where the null subject became less and less used in the language.

2.2 The Synchronic Pattern

Nowadays, the synchronic pattern of BP reflects the diachronic pattern, as we can see in Graph 3. The graph was taken from Duarte's sociolinguistic study from 1995. She conducted an apparent time study with data from 13 adult speakers of BP from Rio de Janeiro, split into three

groups: Group 1 (59 to 74-year-olds), Group 2 (45 to 53-year-olds) and Group 3 (25 to 32-year-olds)[6].

Graph 3: Percentage of definite null subjects in the three persons in BP according to age group (adapted from Duarte, 1995: 48)



As an examination of Graph 2 and Graph 3 suggests, the difference observed between the three groups seems to reflect diachronic changes: younger speakers tend to use the null subject less frequently. This is specially true when we compare Group 1 (oldest speakers) and Group 3 (youngest speakers). Graph 3 shows that the decreasing use of null subjects has affected each grammatical person in the following order, from more to less affected: second person, first person and third person. This is exactly the same pattern we find presented in Graph 2: the second person was more drastically affected by the change in the system of null subjects,

[6] The sample was taken from NURC-RJ project. All participants had at least a college degree. They were interviewed in the 70's, with the exception of the youngest group, whose time when they were recorded is unknown.

followed by the first and third person. In other words, the person that was historically more affected by the change in the verbal paradigm (second person) is the one that exhibits the lowest frequency of null subjects in the synchronic pattern. The person that was historically less affected by the change in the verbal paradigm (third person) is the one that exhibits the highest frequency of null subjects in the synchronic data. The same pattern where the occurrence of null subjects is higher in the third person, followed by the first and second person is also reported by Lira (1982).

2.3 The Status of Definite Null Subjects in BP

Differently than consistent null-subject languages, there are no obligatory contexts for definite null subjects in BP, that is, they occur in the same contexts where overt pronouns occur (Duarte 1995). In consistent null-subject languages, pronouns in embedded clauses with co-referent subjects are generally obligatorily null: overt pronouns are interpreted as having distinct reference than the subject in the higher clause. As we can see in (28), the co-referent embedded subject is obligatorily null in Italian, but optional in BP (29)[7]:

(28) Gianni$_i$ dice che (*lui$_i$) vuole comprare una macchina. *Italian*
Gianni say:3SG that (*he) wants buy:INF a car
'Gianni says that he wants to buy a car.'

(Holmberg 2010a: 91)

[7] In cases involving contrastive focus, the overt pronoun can be co-referent with the subject in the matrix clauses in languages like Italian (Jovović 2020). In BP, contrastive focus is not necessary for co-reference.

(29) O João diz que (ele) quer comprar um carro. *BP*

The John say:3SG that (he) wants buy:INF a car

'John says that he wants to buy a car.'

As we saw, in BP, definite null subjects are possible in embedded clauses when controlled by a subject in a higher finite clause. As already discussed in section 1.3, null subjects in this context also occur in Finnish, another partial null-subject language (the same is true for Marathi , another partial null-subject language (Holmberg and Sheehan 2010)). Other contexts in which definite null subjects are possible in BP are: (i) null subjects in matrix clauses when the antecedent can be easily identified by the discourse; (ii) null subjects in sentences in which the antecedent is absent in the discourse, but salient in the pragmatic context (i.e., it is identified by the physical presence of the referent) (Simões 1999; Magalhães 2006). In what follows, I discuss null subjects in these two particular contexts.

The context in (i) commonly involves subjectless replies (Modesto 2000b), as can be seen in the example below:

(30) Q: O que o João fez? *BP*

What that the John did:3SG

'What did John do?'

A: *e* dormiu o dia inteiro.

slept:3SG.PAST the day whole

'He slept all day long.'

Importantly, subjectless replies are not derived by the same mechanism spelled out in section 1.3. A proposal is offered by Modesto (2000b), for whom the null subject in (30) is derived by topic-deletion. That is, the definite null subject is possible in (30), according to his proposal, because it refers to 'João', a topic phrase which was deleted in the answer in (30). I refer the reader to Modesto (2000b) for the details of his proposal.

In this thesis, subjectless replies will not be discussed. It is worth noticing, though, that they also occur in non-null-subject languages, although they might be derived through different mechanisms than subjectless replies in BP (see Holmberg 2003 for an analysis of these structures in non-null-subject languages). The Swedish example below illustrates subjectless replies in non-null-subject languages:

(31) Q: Vad gör Johan? *Swedish*
what does Johan
'What's Johan doing?'

A: Sover.
sleeps
'He's sleeping.'

(Holmberg 2010a: 90)

In the context in (ii), the null subject is identified by the physical presence of the referent, as the example in (32):

(32) [context: a cup falls to the floor and breaks into pieces]

e quebrou!

broke:3SG.PAST

'The cup broke!'

As pointed out by Simões (1999), cases like (32) are marked in BP, given the fact that the sentence requires a "here and now" context (see discussion in Chapter 6). Null subjects recovered by the physical presence of the referent are not the object of study of this thesis. For our purposes, it is enough to notice that in consistent null-subject languages, null subjects occur even when their referent is absent in the physical context. Due to this reason, I assume that the null subject in (32) is not derived by the presence of D in T.

In sum, we saw in this section that there are three types of definite null subjects in BP, none of which involves the presence of D in T: (i) embedded null subjects co-referent with a subject in the matrix clause; (ii) null subjects which appear in replies and (iii) null subjects identified by the physical presence of the referent.

3 Generic Null Subjects

In this section, I distinguish between different types of generic null subjects and specify which ones are relevant for impersonal constructions which I investigate in this thesis.

According to the typology presented in Holmberg and Phimsawat (2015), non-definite generic subjects can be *inclusive*, *quasi-inclusive* or *exclusive*. The *inclusive* pronoun refers to people in general including the speaker and the addressee[8]. *Quasi-inclusive* pronoun refers to people in general including the speaker but not necessarily the addressee. *Exclusive* pronouns refer to people in general excluding the speaker and the addressee. These pronouns are overt in some languages and null in others. Constructions that have these types of pronoun are referred to as *generic impersonal constructions*. The sentences in (33) illustrate each one of the readings for the null subject in impersonal constructions in English:

(33) a. **One** should sleep at night. *inclusive*

= People, including you or me should sleep at night.

b. **We** like coffee in Brazil. *quasi-inclusive*

= People, including me but not necessarily you like coffee in Brazil.

c. **They** speak three different languages in that house. *exclusive*

= People, but not you or me speak three different languages in that house.

[8] As Moltmann (2007:281) points out, generic inclusive pronouns such as *one* do not range exactly over people, but rather conscious beings, given the acceptability of the sentence below (Safir 2000):
(i) If one is a Martian, one is not susceptible to human disease.

By analyzing data across different types of null subject languages, Italian (consistent)[9], Thai (radical) and Finnish (partial), Holmberg and Phimsawat (2015) observe that the inclusive, quasi-exclusive and exclusive impersonal null subject is realized in different ways across these languages[10]:

Table 3: Generic pronouns across different null subject languages

Generic Pronoun	**Italian (Consistent)**	**Thai (Radical)**	**Finnish (Partial)**
Inclusive (one)	Overt *(si)*	Null	Null
Quasi-inclusive (we)	Null or overt	Overt	Null or overt
Exclusive (they)	Null	Null or overt	?

[9] I presented the following sentences to an adult speaker of another consistent null-subject language, European Portuguese (46-years-old, from Faro, Algarve), asking if the pronoun could refer to (a) specific individual(s) and/or people in general, including/excluding speaker and/or addressee. His judgments were equivalent to what we see in Table 3 for Italian. Where * is seen, it indicates that the non-definite reading is not possible, but the definite can be:

(1)a. Em Lisboa não **se** pode beber mais chá. (In Lisbon one can't drink tea anymore, *overt*)

(1)b. *Em Lisboa não pode beber mais chá. (In Lisbon one can't drink tea anymore, *null*)

(2)a. De acordo com o primeiro ministro, **nós** temos de ser mais produtivos. (According to the PM, we need to be more productive, *overt*)

(2)b. De acordo com o primeiro ministro, temos de ser mais produtivos. (According to the PM, we need to be more productive, *null*)

(3)a. *Em Lisboa, **eles** não bebem mais chá. (In Lisbon they don't drink tea anymore, *overt*)

(3)b. Em Lisboa, não bebem mais chá. (In Lisbon they don't drink tea anymore, *null*)

The sentence in (1b) is impossible with the inclusive and the exclusive generic reading of the pronoun. This observation will be relevant for the discussion in section 3.3.

[10] Table 3 implies a prediction that all consistent null-subjects should pattern like Italian, all radical pro-drop languages should pattern like Thai and all partial null-subject languages should pattern like Finnish (with the exception of '?' for exclusive 'they'). BP, though, does not pattern like Finnish, as discussed below.

Although BP is also considered to be a partial null-subject language (Holmberg et al. 2009), it patterns different from Finnish in its expression of generic pronouns. The inclusive reading of the subject could be expressed either by a null subject or by an overt pronoun (*se* or *você*).

(34)a. Em Lisboa não **se** pode beber mais chá.
In Lisbon not SE can drink:INF more tea
'In Lisbon one can't drink tea anymore.'
= People, including you or me can't drink tea in Lisbon anymore.

b. Em Lisboa *e* não pode beber mais chá.
In Lisbon not can drink:INF more tea
'In Lisbon one can't drink tea anymore.'
= People, including you or me can't drink tea in Lisbon anymore.

c. Em Lisboa ***você*** não pode beber mais chá.
In Lisbon you not can drink:INF more tea
'In Lisbon one can't drink tea anymore.'
= People, including you or me can't drink tea in Lisbon anymore.

As Holmberg (2010: 93) points out, "of the partial null-subject languages, some accept null exclusive as well as inclusive generic pronouns (BP), other just inclusive (Finnish)". The

issue with the exclusive reading in Finnish is that these constructions have an impersonal verb form, identified also as a passive (Blevins 2003; Manninen and Nelson 2004). According to Holmberg and Phimsawat (2015), it is controversial whether there is a null subject in a sentence like (35) with the exclusive reading in Finnish, although there is no overt subject either. That is why the possibility of the exclusive reading in Finnish appears as "?" in the table above.

(35) Intiassa puhutaan monta eri kieltä.
India:INE speak:IMP many different language
'They speak many different languages in India.'
= People, but not you or me speak different languages in India.

An example of the exclusive reading of the null pronoun in BP is given in (36) below:

(36) Em Lisboa *e* não bebem mais chá.
In Lisbon not drink:3PL more tea
'In Lisbon they don't drink tea anymore.'
= People, but not you or me don't drink tea in Lisbon anymore.

An overt exclusive pronoun is also possible in BP, which makes it different from both consistent null-subject languages and Finnish. In the example below, the pronoun *eles* (they) refers to people in general (from Lisbon) and it excludes the speaker and the addressee.

(37) Em Lisboa **eles** não bebem mais chá.

In Lisbon they not drink:3PL more tea

'In Lisbon they don't drink tea anymore.'

= People, but not you or me don't drink tea in Lisbon anymore.

The overt quasi-inclusive pronoun is realized by *a gente* in BP. Before its grammaticalization that started in the 16th century (Lopes 2001), *a gente* was an expression that meant "the people" (this use is still possible nowadays). Therefore, it makes sense that *a gente* can have a generic reading, referring to "people in general". The example below was taken from a play by França Júnior, written in the second half of the 19th century. Already by this time, *a gente* had the quasi-inclusive reading, including the speaker, but not necessarily the addressee (quasi-inclusive):

(38) A prima Maricota disse-me que era uma coisa de pôr **a gente** de

The cousin Maricota told-CL:1SG that was a thing of put us of

queixo caído

jaw dropped

'The cousin Maricota told me that it was something to make our jaw drop.'

(Júnior 1982: Scene VII)

The example in (39) includes the pronoun *a gente* in the subject position with a quasi-inclusive reading (in (38) *a gente* is in the object position). The speaker is talking about his

childhood and *a gente* refers to children in general at that time, including the speaker when he was a child:

(39) **A gente**, por exemplo, era guri, **a gente** vinha do colégio e coisa,
We for example were children we come:IMP of.the school and thing
entrava uma pessoa idosa
enter:PST a person elderly
'We, for example, were children, we were coming back from school and so on, an elderly person would come into [the bus]…'

(VARSUL corpus, Zilles 2002)

The null quasi-inclusive pronoun is not possible in BP with the third person singular agreement marker (the one used by *a gente*). When a sentence is uttered out of the blue with the third person singular agreement mark, it has the inclusive reading in BP (*one*). As the first plural inflection is rarely used nowadays in spoken BP, we could say that a sentence like (41), possible in EP with the quasi-inclusive reading and definite reading, would not be productive in BP:

(40) De acordo com o primeiro ministro, temos de ser mais produtivos.
According with the primer minister should of be more productive
'According to the PM, we need to be more productive.'

The table below summarizes how generic pronouns are realized in BP and EP (for comparison):

Table 4: Generic pronouns in Brazilian Portuguese and European Portuguese

Generic Pronoun	**Brazilian Portuguese (Partial)**	**European Portuguese (Consistent)**
Inclusive (one)	Null or overt	Overt
Quasi-inclusive (we)	Overt	Null or overt
Exclusive (they)	Null or overt	Null

In the following subsections, the availability or nonavailability of each one of these readings in BP is explained in more details. Additionally, the difference between generic and existential sentences is explained in the subsection 3.4.

3.1 Inclusive Reading

In BP, the inclusive reading of the null subject can be realized only with the 3rd person singular agreeing marker. The inclusive reading is impossible with the 3rd person plural agreeing marker: with the third person plural marker on the verb, the null subject has a exclusive reading.

(41)a. Nessa biblioteca *e* não pode estudar.
In.this library not can:3SG study:INF
'In this library one cannot study.'
= People, including you or me can't study in this library.

b. Nessa biblioteca *e* não podem estudar.

In.this library not can:3PL study:INF

'In this library they cannot study.'

≠ People, including you or me can't study in this library.

= People, but not you or me can't study in this library.

In addition to the null inclusive subjects, BP can have overt inclusive subjects, realized by *você*[11] and by *se*, as seen below:

(42) Nessa biblioteca você não pode estudar.

In.this library you not can:3SG study:INF

'In this library you cannot study.'

= People, including you or me can't study in this library.

(43) Nessa biblioteca não se pode estudar.

In.this library not SE can:3SG study:INF

'In this library one cannot study.'

= People, including you or me can't study in this library.

[11] As discussed in section 2.1, although *você* is a second-person pronoun it combines with a unmarked third person singular form.

Consistent null-subject languages do not have a null inclusive generic pronoun. They rely on overt strategies. Romance and some Slavic languages make use of cognates of the overt reflexive *se* pronoun. Hausa makes use of a special impersonal pronoun *á* (Jaggar 2001). Moroccan Arabic and Greek use generic *you* with 2nd person singular agreement.

A minimalist proposal that explains the lack of incorporated inclusive generic subjects in consistent null-subject languages, such as Italian and EP, and the use of a null pronoun in partial null-subject languages, such as BP, is found in Holmberg (2010a). As explained in section 1.2, consistent null-subject languages include a definiteness feature in T, while partial null subjects do not, as seen in the contrast between (21) and (23), repeated here in (44) and (45), respectively:

(44)a. [T, D, uφ, NOM] [*v*P [3SG, uCase] *v*...] → *CNSL* (definite pronoun)

b. [T, D, 3SG, NOM] [*v*P [3SG, NOM] *v*...] →

c. [T, D, 3SG, NOM] [*v*P [~~3SG, NOM]~~ *v*...]

(45)a. [T, uφ, NOM] [*v*P [3SG, uCase] *v*...] → *PNSL* (generic inclusive pronoun)

b. [T, 3SG, NOM] [*v*P [3SG, NOM] *v*...] →

c. [T, 3SG, NOM] [*v*P [~~3SG, NOM]~~ *v*...]

The structure in (44) derives a definite null subject, while the structure in (45) derives a generic null subject, due to the absence of a D-feature in T. In consistent null-subject languages, generic null subjects cannot be null: as T in these languages have a D-feature, a null pronoun will

always have the definite reading. The clitic pronoun *se/si* is therefore inserted to derive the generic reading of the pronoun in languages like European Portuguese and Italian.

It seems that nothing prevents that a language which uses generic null subjects can also use overt strategies to express the inclusive generic pronouns, given the pattern found in BP. As explained above, two overt strategies to derive the inclusive generic reading are available in BP: the use of *você* and *se*. Regarding inclusive *se*, as we will see in detail in Chapter 4 and 5, it is becoming less and less productive in BP and it is not acquired early by children. *Você* is a common strategy to express the generic reading, though (Assis 2017).

Holmberg (2010a) is not explicit about the structure of overt generic subjects. According to Egerland (2003) and Fenger (2017), there are two types of overt generic subjects: overt generic subjects with φ-features (46a) and without it (46b):

(46)a. $[[_{\varphi P}\ \varphi\ [_{NP}\ N]]]$ *imp-φ*

b. $[_{NP}\ N]$ *imp-N*

According to Fenger (2017), imp-φ pronouns do not allow any existential reading (i.e., referring to *someone* or *somebody*), while imp-N pronouns do. The φ-features restrict the interpretation of a pronoun, making it compatible only to certain persons in the discourse (Egerland 2003). This yields the different generic readings: inclusive, quasi-inclusive and exclusive reading of the pronoun. When a pronoun lacks φ-features (i.e., imp-N pronoun), it does not have any interpretative restriction and the pronoun will become compatible with either the

generic or the existential reading. In other words, while imp-φ pronouns can only have the genetic reading, imp-N pronoun can have both the existential and the generic reading.

According to Cinque (1988), in the existential reading, a single individual can satisfy the reference of the pronoun. Besides, this reading is compatible with a specific time reference. Below, I test whether the existential reading is compatible with *você* and *se.*

(47) Intended: 'Somebody has played soccer terribly yesterday in São Paulo, but I don't know who.

a. Ontem em São Paulo se jogou futebol muito mal. *BP*
Yesterday in São Paulo SE played soccer very bad
'Yesterday somebody in São Paulo played soccer very terribly'

b. *Ontem em São Paulo você jogou futebol muito mal.
Yesterday in São Paulo you played soccer very bad
'Yesterday somebody in São Paulo played soccer very terribly.'

We see in (47) that, given the appropriate context, *se* can have the existential reading, in contrast with *você*, which cannot have the existential reading even when the context is manipulated to favor this reading. We can conclude that *se* is an imp-N pronoun, due to the availability of the existential reading when the context favors it. *Se* can also have the generic

reading, as discussed in section 3.1. Regarding *você*, the lack of existential reading of this forms indicates that it is an imp-φ pronoun.

If *se* is an imp-N pronoun, lacking φ-features, how does T get its unvalued φ-features valued? According to Chomsky's (2001) theory of agreement, subject-verb agreement is encoded by unvalued φ-features in T. These features are valued by the subject DP. If unvalued φ-features are not assigned a value, the derivation crashes at PF, as the finite verb cannot be spelled out. A possible solution for this problem is that T has inherently valued φ-features (i.e., 3SG), that is, default agreement, since *se* cannot value the φ-features of T if *se* is an imp-N.[12]

I conclude this section by anticipating that the generic pronouns studied in the survey to be reported in Chapter 3 and in the experiments to be reported in Chapter 4 and 5, respectively, have the inclusive reading.

3.2 Quasi-inclusive Reading

The quasi-inclusive pronoun refers to people in general including the speaker but not necessarily the addressee. As can be seen by a comparison between Table 3 and Table 4, quasi-inclusive pronouns can be optionally null in Finnish, a partial null-subject language, but they cannot be null in BP, at least not as a counterpart of the overt pronoun *a gente* that has unspecified agreement on the verb. In Finnish, the verb that appears with a null quasi-inclusive pronoun has first person plural agreement marker:

---

[12] Default agreement is assumed in weather expressions in Finnish, for example, where there is no element to values T' φ-features (Holmberg 2010b: 206):

(i) Nyt sataa.
Now rain:3SG
'Now it rains.'

(48) (Me) syömme Soumessa paljon savukalaa.

We eat:1PL Finland:INE much smoke.fish

'We eat a lot of smoked fish in Finland.'

(Holmberg and Phimsawat 2015: 57)

In EP, the pronoun *nós* coexists with the pronoun *a gente.* The agreement marker for the first person plural was largely preserved in EP. While in BP the form *a gente* occurs in 80% of the time (Zilles 2007), in EP *nós* is the preferred form, occurring in 82% of the time (Vianna 2012). In EP, the quasi-inclusive generic pronoun can be null:

(49) … porque *e* estamos habituados na construção civil é construir prédios.

because are used in.the construction civil is build:INF buildings

'Because what we use to do in the civil construction is to build buildings.'

(Assis 2017: 118)

To the extent that BP maintains *nós* and the first person plural agreement marker, the sentence (49) would be possible in the language, although, it is not productive. Rather than this form, the way to express the quasi-inclusive reading in BP is by using overt *a gente*.

In Thai, where the verb does not carry an agreement marker, the quasi-inclusive pronoun can be null, but only in an embedded sentence where it is controlled or bound by an overt matrix subject.

(50) raw kin cee nay d~~u~~an tùlaakhɔm lăŋ *e* thamboonsàjbàat.

we have veg.food in month October after offer food to monk

'We have vegetarian food in October after offering food to monks.'

(Holmberg and Phimsawat 2015: 64)

The sentence equivalent to (50) is good in BP if *a gente* controls the null embedded subject:

(51) A gente tem comida vegetariana em Outubro depois de *e* oferecer

We have:3SG food vegetarian in October after of offer:INF

comida aos monges.

food to.the monks

'We have vegetarian food in October after offering food to the monks.'

For our purposes, it is enough to make clear that a sentence with the null subject and the verb carrying the third person singular agreement marker does not have the quasi-inclusive reading in BP if it is not controlled or bound by the overt pronoun *a gente*. Since the null subject in the test sentences from the survey to be reported in Chapter 3 and the experiments to be reported in Chapters 4 and 5 happen in matrix clauses, these null subjects do not have the quasi-inclusive reading.

As the quasi-inclusive reading of the null pronoun is not relevant for the present study, the precise mechanism that licenses its reading will not be discussed (see Phimsawat 2011; Holmberg and Phimsawat 2015).

3.3 Exclusive Reading

As for null pronouns with an exclusive reading, which exclude the speaker and the addressee, BP can have the null or overt pronoun, while EP only has the null counterpart (see footnote 9). In BP, the optional null subject with the exclusive reading is realized with the third person plural agreeing marker on the verb (52). If this pronoun is realized with a verb carrying the third person singular agreeing marker, it tends to receive an existential reading (equivalent to 'someone'), rather than the generic reading.[13] In EP, exclusive generic null subjects must be realized with the plural agreeing marker as well, as shown in (53a) (Duarte and Figueiredo da Silva 2016). The sentence in (53b) is only possible with the definite reading in EP.[14]

(52)a. Em São Paulo (eles) comem bem. *BP*

In São Paulo eat:3PL well

'They eat well in São Paulo.'

= People, but not you or me eat well in São Paulo.

---

[13] As will be discussed in section 3.4 and 5, it is a hypothesis raised in this chapter that the third person singular null pronoun can be interpreted as generic in a sentence like (52b) only if it has a genericity marker (e.g., a deontic modal or the clitic *se*). This generic reading of the third person null pronoun is inclusive. Without the genericity marker this pronoun is interpreted as existential.

[14] The judgements in (50) were confirmed by a native speaker of EP (the same speaker who judged the sentences in footnote 9). Duarte and Figueiredo da Silva (2016) also discuss that EP lacks generic null subjects in the third person singular form, which is in accordance with Holmberg's (2010a) proposal.

(52)b. Em São Paulo (ele) come bem. [*generic / existential]

In São Paulo eat:3SG well

'They eat well in São Paulo.'

= Someone in São Paulo eats well.

≠ People, but not you or me eat well in São Paulo.

≠ People, including you or me eat well in São Paulo.

(53)a. Dizem que o governo vai aumentar outra vez os impostos. *EP*

Say:3PL that the government will raise:INF again the taxes

'They say that the government will raise taxes again.'

= People, but not you or me say the government will raise taxes again.

(Duarte and Figueiredo da Silva 2016: 242)

b. *Diz que o governo vai aumentar outra vez os impostos.

Say:3SG that the government will raise:INF again the taxes

'They say that the government will raise taxes again.'

= People, but not you or me say the government will raise taxes again.

The fact that consistent null-subject languages allow exclusive generic null subjects leads Holmberg and Phimsawat (2015) to propose that 3PL exclusive pronouns are not φP pronouns licensed in the same way as inclusive generic null subjects (see section 3.1). That is so because

the presence of D in T in consistent null subject languages makes a φP pronoun obligatorily definite. Therefore, the exclusive 3PL generic pronoun must be licensed by a different mechanism than the inclusive generic pronoun.

Following Brody (2013), Holmberg and Phimsawat (2015) adopt the proposal that exclusive generic null pronouns are licensed by a covert noun 'people' (i.e., 'silent people' in Brody's term). This covert noun is an optional null Topic, which, according to Holmberg and Phimsawat (2015), is preceded by a scene-setting locative or a temporal PP as in the presentation below:

(54) [$_{CP}$ [nesta aldeia] [$_{TopP}$ (as pessoas)$_i$ Top [$_{TP}$ e_i tomam chá]]] *EP*

in.this village the.people drink:3PL tea

'In this village people drink tea.'

= People, but not you or me drink tea in this village.

According to Holmberg and Phimsawat (2015), the null pronoun in (54) is not a φP pronoun, but a bound/controlled null subject.

As for the requirement that the null Topic in sentences like (54) should be preceded by a locative PP or temporal adverbial, this does not seem to be obligatory, given the grammaticality of sentences like (53a) in EP which lack an adverbial. It is possible, though, that (53a) is interpreted as having a null adverbial.

Since the null subject in the test sentences from the survey to be reported in Chapter 3 and the experiments to be reported in Chapters 4 and 5 have a third person singular pronoun rather than third person plural pronoun, these null subjects do not have the exclusive reading.

3.4 Existential Reading vs. Generic Reading

Carvalho (2018, 2019) argues that the null subject in impersonal sentences in BP can have two readings: generic or existential. The existential null subject can be paraphrased as 'someone', 'somebody', that is, they indicate the participation of a non-specified individual in an event. In generic impersonal structures, in contrast, everyone involved in the event described by the sentence is a potential participant.

According to the author, generic and existential impersonal structures have two different pronouns, rather than reflecting the availability of two different readings for the same pronoun. The existential pronoun, according to Carvalho (2019), does not have a person feature, since this pronoun is unable to license anaphors and body part nouns interpreted as inalienable possession, which must be bound by the pronoun[15]:

(55)a. *Na feira *e* não escuta a si mesmo.
In.the street.market not listen:3SG to REFL self
'Someone does not listen to himself in the street market.'

b. *Na escola *e* levanta a mão quando tá com dúvida.

[15] According to Carvalho (2018), the null existential pronoun is marked with number, since it refers to an element in the singular form.

In.the school raise:3SG the hand when is with doubt

'Someone raises their hand at school whenever they have a doubt.'

Adapted from Carvalho (2019: 57)

The generic pronoun, in contrast, has a person φ-feature, as shown by the sentences below. In (56a), the generic null subject binds the body part noun interpreted as inalienable possession. In (56b), the subject licenses an anaphor.

(56)a. Como *e* enxuga a mão nesse aparelho?

How dry:3SG the hand in.this device

'How does one dry one's hand in this device?'

b. Como *e* se enxuga sem toalha?

How SELF dry:3SG without towel

'How does one dry oneself off without a towel?'

(Carvalho 2018: 92)

Besides differing in their feature make-up, there are other characteristics that tease apart existential and generic impersonals. According to Carvalho (2018), existential impersonals are habitual, in the sense that they describe an event that happens habitually in which an unspecified agent takes part (Carvalho 2018; 2019). In contrast, generic impersonals have a rule-like reading

in which the null subject refers to any agent subject to a certain rule or instruction: the action described in the generic sentence can be habitual or not. In the *"how*" sentences in (56), for example, the generic pronoun is any agent who must follow the instructions on how to dry hands in a certain device (56a) and how to dry yourself off without a towel (56b).

It should be pointed out, though, that the observation of Carvalho (2018; 2019) that the existential reading is "habitual" does not seem to be in agreement with what other authors say. Carvalho also gives examples of sentences in the present tense and imperfective aspect claiming that they have an existential reading. This is in disagreement for example with Delfitto (2006). Delfitto (2006: 242-243) discusses that the sentences in the imperfective aspect cannot be interpreted as existentially in English, as shown in (57a). While sentences in the past tense (hence in the perfective aspect) are ambiguous between the existential and generic reading, as shown in (57b). Therefore, Carvalho's (2018; 2019) claim that sentences like (58) are acceptable under the existential reading needs to be evaluated.[16]

(57)a. Firemen extinguish the fire. (generic/*existential)

b. Firemen extinguished the fire. (generic/existential)

Delfitto (2006: 242-243)

(58) Naquela loja vende sofá. (*generic/existential)

In.that store sell:3SG sofa

[16] My judgement of (58), before knowing Carvalho's work, is that this sentence is simply ungrammatical. This is so because the tense and aspect favors the generic reading, while there is no overt marker of genericity in this sentence (e.g., a deontic modal), as seems to be required in BP (see section 5).

'Someone sells sofa in that store.'

Carvalho (2019: 47)

Another difference between existential and generic impersonals, as will be shown to be relevant for the discussion in section 4, is that existential sentences require a locative as an event argument. Generic impersonals, on the other hand, do not require a locative (as seen in (56)). The reason why a locative is required in existential impersonals lies on the fact that *existential pronouns are only licensed as an external argument of a stage-level transitive vP.* The locative is an obligatory event argument only in stage-level predicates (Diesing 1992 and Kratzer 1995)[17].

As seen by the contrast in (59), the existential pronoun can only occur with stage-level predicates, not with individual-level ones. As pointed out by Carvalho (2018), generic pronoun can occur with both types of predicates (60), and therefore, they do not require a locative as an event argument (as we see in (59)), although some generic sentences might have a locative (as in (60) when a rule applies to a specific place).

---

[17] Diesing (1992) and Kratzer (1995) capture the difference between stage-level and individual-level predicates in terms of an event variable. If we look at the examples below from Kratzer (1995: 136), while the stage-level predicate *hit* has an event argument for spatiotemporal location in their theta-grid, the individual-level predicate *know* lacks it:

(1) a. hit <location, agent, theme> *stage-level*
b. know <experiencer, theme> *individual-level*

As shown in the contrast below, it is a general property of stage-level predicates to allow modification by locatives and temporal adverbials, while individual-level predicates do not allow this modification (Chierchia 1995). Therefore, 'location' is specified in the theta-grid of a stage-level predicate (2), but not in the theta-grid of an individual-level predicate (3).

(2) John hit Peter in the car. *stage-level*
(3) *John knows French in the car. *individual-level*

(59)a. Nessa escola *e* ensina matemática. *stage-level* *BP*

In.this school teach:3SG math

'In this school someone teaches math.'

b. *Nessa escola *e* sabe matemática. *individual-level*

In.this school know:3SG math

'In this school someone knows math.'

(Carvalho 2018: 85)

(60)a. Nessa escola *e* tem que ensinar matemática. *stage-level*

In.this school have that teach:INF math

'In this school one has to teach math.'

b. Nessa escola *e* tem que saber matemática. *individual-level*

In.this school have that know:INF math

'In this school one has to know math.'

While the stage-level predicate in (59a) and (60a) describes a specific, temporary property (or "stage") of the school (e.g., they teach math there right now, but they may not teach it anymore in the future), the sentences in (59b) and (60b) describe a property that the school, or those that are part of the school have that at least has the tendency to be stable.

For our purposes, it is important to know that the sentences used in the experiment to be reported in Chapter 4 and 5 were generic sentences with an inclusive reading, rather than sentences with the existential reading. That is so because (i) all the sentences were presented with a lawlike background, as will be discussed in more detail in section 5, and (ii) they had the third person singular agreeing marker on the verb. Some of the sentences employed in the survey to be reported in Chapter 3 have the existential reading, though.

4 Impersonals and the EPP feature

I will discuss in this section Holmberg's (2010a,b) claim that generic impersonal structures require an adverbial or any element that can check the EPP (Extended Projection Principle) in order to be well-formed. The EPP was originally the requirement that the subject in finite clauses should be filled in [Spec, TP] (Chomsky 1981). In the Minimalist Program, this condition was replaced by the claim that T has an uninterpretable EPP-feature which requires the presence of XP in the specifier of T in order to delete this uninterpretable feature.

According to Holmberg (2010b), generic φP pronouns are unable to satisfy the EPP. This is so because φP pronouns are the non-head member of an argument chain headed by T located in [Spec, vP]. Grammatical operations applying to a chain cannot apply to the non-head member of a chain, but only to the head of a chain. As φP pronouns are non-head members of a chain, they are not accessible to the EPP.[18] According to the author, in structures with generic null subjects, the EPP cannot be checked by the goal of T's probing: another element has to merge with TP to check the EPP. According to Holmberg (2010a), this means that the EPP is φ-

[18] Recall that in consistent null-subject languages, an Aboutness-shift topic checks the EPP in Holmberg's (2010a) proposal, not the φP pronoun, as discussed in section 1.3.

independent in partial null-subject languages. In Finnish, an overt expletive can be inserted for this purpose or the EPP can be checked by moving a non-subject category, such as an object, a locative or a temporal adverbial to the preverbal position (Holmberg 2010b), as in (61). In (62), the structure of (61b) is given: as can be seen, the locative *tässä* merges with TP and deletes the uninterpretable EPP-feature of T.

(61)a. *Istuu mukavasti tässä. *Finnish*
sit:3SG comfortably here
'One sits comfortably here.'

b. Tässä istuu mukavasti.
here sit:3SG comfortably
'One can sit comfortably here.'

c. Sitä istuu mukavasti tässä.
EXP sit:3SG comfortably here
'One can sit comfortably here.'

(Holmberg 2010b: 204)

(62) $[_{TP}$ tässä $[_{T'}$ istuu + $T_{3SG,\ \cancel{EPP}}$ $[_{VP}$ φP_{3SG} <istuu> mukavasti <tässä>]]]]

(Holmberg 2010a: 105)

Holmberg (2010b) observes that there are exceptions to the EPP in Finnish, as the examples below in (63) show (data from Laitinen 2006). Holmberg (2010b) has no explanation for these exceptions, where the generic null subject is possible without any other element to check the EPP. He claims, though, that minimal pairs such as (61a) and (61b) are pervasive in Finnish.

(63)a. Täytyy harjata hampaat. *Finnish*

Must brush teeth

'One/I/you must brush one's/my/your teeth.'

b. Ei saa tulla sisään kengät jalassa.

not may come in shoes foot:INE

'You may not come in with shoes on.'

c. Jos/kun *e* ostaa auton Saksassa,…

If/when buy:3SG car:ACC Germany:INE

'If/when you buy a car in Germany,…'

d. Sitä on oltava hyvin tarkka jos *e* ostaa auton Saksassa.

EXP is be-PRC very careful if buys car Germany:INE

'You have to be careful if you buy a car in Germany.'

(Holmberg 2010b: 212)

Regarding BP, the empirical results to be reported in Chapter 3 investigating the grammar of adults seem to suggest that the EPP is not relevant for the well-formedness of sentences with the generic null pronoun. As we will see in detail in Chapter 3, the results obtained in a survey I conducted with BP-speakers can be explained if the only requirement in generic impersonal structures is to have an overt genericity maker (this can be an overt generic pronoun or a deontic modal, as discussed in the next section of this chapter).

The literature commonly refers to the role of fronted adverbials in checking the EPP in sentences with non-referential null subjects in BP (Rodrigues 2004; Pilati, Nave and Salles 2017; Holmberg et al. 2009). The contrast in (64) is usually interpreted as indicating that the locative is required in [Spec, TP] when the null subject has a generic reading in BP (64a). In the absence of a locative, the null subject in (64b) can only have the definite reading (the null subject here is controlled by an argument in the higher clause), since the copy of the controlled subject occurs in [Spec, TP] in this sentence, it can check the EPP.

(64)a. João me contou que na praia *e*. vende cachorro quente. *BP*

John me told that at.the beach sell:3SG dog hot

'John told me that hot dogs are sold at the beach.'

≠ 'John told me that he sells hot dogs at the beach.'

b. João$_i$ me contou que e_i vende cachorro quente na praia.

John me told that sell:3SG dog hot at.the beach

'John told me that he sells hot dogs at the beach.'

≠ 'John told me that hot dogs are sold at the beach.'

(Rodrigues 2004: 142)

Carvalho (2018, 2019) defends the idea that in sentences like (64a), the null subject is existential instead of generic (i.e., *someone* sells hot dogs at the beach). For the author, the locative occurs in a sentence like (64a) because it is an event argument of the stage-level predicate present in every impersonal sentence with an existential reading.

As observed in section 3.4, some sentences with generic null subject in BP have a locative in order to specify the place to where a rule applies. As the locative does not need to be fronted, as shown by the fact that both (65a) and (65b) are possible, the locative does not have the role to check the EPP.

(65)a. Tem que comer banana nessa escola. *BP*

Have to eat:INF banana in.this school

'One has to eat bananas in this school.'

b. Nessa escola tem que comer banana.

In.this school have to eat:INF banana

'One has to eat bananas in this school.'

Rather than claiming that the generic null subject is able to check the EPP, I follow Bošković's (2002) proposal that the EPP, as traditionally formulated, does not hold and where it seems to hold, its effects follow from independent reasons. For the particular case of sentences with generic null subjects, I suggest that the EPP effect follows from a semantic requirement to have the place where the rule-like statement holds specified whenever it is necessary. In other cases, where the rule holds whatever place an individual is at, the locative is unnecessary, as in the sentence below:

(66) Não pode largar a escola. *BP*
Not can:3SG leave:INF the school
'One cannot drop out of school.'

It is generally advised that one should not drop out of school and this holds in any place we can imagine. For this reason, the locative becomes unnecessary in the sentence. There are also cases in which the locative is implicit in the context and does not need to be overt. For example, the sentence in (67) is grammatical: if a speaker utters the sentence in (67) in a store, it is not necessary to say "in this store", since the place where the rule holds is clear for the addressee.

(67) (Nessa loja) não pode vender sofá. *BP*
(In.this store) not can sell:3SG sofa
'(In this store) one cannot sell sofas.'

Although I do not adopt Holmberg's (2010a,b) idea that there is an EPP feature in T to be checked by an element other than the subject in partial null-subject languages, I do adopt his position that the generic null subject is a ϕP pronoun located in [Spec, vP]. Very often, an element other than the goal of T's probing (e.g., a locative) will be found in [Spec, TP] in structures with generic null subjects, but this is not always the case: sometimes, [Spec, TP] seems to be empty.

In Chapter 6, I will adopt Holmberg's (2010a) terminology that BP and other partial null-subject languages have ϕ-independent EPP, although EPP here is used pretheoretically (i.e., I do not assume that there is, in fact, an EPP feature in T to be checked). Holmberg (2010a) formulated the ϕ-dependent EPP parameter to describe whether the DP or pronoun which values T's features checks the EPP (in this case, the language is ϕ-dependent) or another category does (in this case, the language is ϕ-independent). Although I assume that in BP the reason why a locative can appear in [Spec, TP] in generic impersonals is not because T has an EPP feature, I adopt Holmberg's terminology with a slightly different meaning. I assume that ϕ-independent languages are languages in which [Spec, TP] can be filled by an element other than the DP or the pronoun which values T's features, as in (68):

(68)a. Φ-dependent languages: [Spec, TP] is filled by the DP or pronoun which values T's features.

b. Φ-independent languages : [Spec, TP] can be filled by an element which does not value T's features.

As we will see in detail in Chapter 6, consistent null-subject languages and non-null-subject languages are ϕ-dependent, while partial null-subject languages are ϕ-independent. This concept will be important to explain the acquisition of different null-subject languages (Chapter 6).

5 Overt Expressions of Genericity in Impersonal Constructions

In this section, I will propose that deontic modals and the impersonal clitic *se* are overt expression of genericity. Generic impersonal sentences without an explicit expression of genericity do not seem to be completely well-formed in BP (see Chapter 3). *Se* and the deontic modal would have the function to express the generic reading of impersonals.

The impersonal sentences that were part of Experiment 1 (Chapter 4) and 2 (Chapter 5) were all about rules that happen in a certain school. Sentences with a generic pronoun, as mentioned in section 3.4, talk about a rule in relation to a certain event. We can call these structures *characterizing sentences,* as opposed to *particular sentences*. While particular sentences express statements about particular events and facts, characterizing sentences express general laws or generalizations (Krifka et al. 1995).

Krifka et al. (1995) list some tests that can be used to distinguish characterizing sentences from particular sentences. The first test consists in combining a sentence with an adverb like *usually, typically, often*, *always*, etc. If the original sentence is little affected by the addition of one of these adverbs, then the original sentence is a characterizing sentence. Particular sentences, in contrast, would suffer drastic changes: the sentence will change from a report of a specific event or fact to a generalization. This can be seen by the contrast between (69) and (70):

(69)a. A lion stood in front of my tent. *Particular*

b. A lion usually stood in front of my tent.

(70)a. A lion has a bushy tail. *Characterizing*

b. A lion usually has a bushy tail.

(Krifka et al. 1995: 9)

A second test to help distinguish between particular and characterizing sentences takes into account the fact that characterizing sentences do not express accidental properties, but properties that are essential. For example, the sentences in (71) express an essential property of madrigals: they are polyphonic, therefore, they can only be understood as characterizing sentences. The sentences in (72), by contrast, describe an accidental property of madrigals. Notice that (71c), for having an indefinite singular NP, can only be understood as a characterizing sentence[19], while at the same time it is describing the accidental property of madrigals being popular: therefore, the sentence in (72c) is bad with the characterizing reading.

(71)a. The madrigal is polyphonic.

b. Madrigals are polyphonic.

c. A madrigal is polyphonic.

[19] Indefinite NPs are not considered generic by themselves, though. For discussion, see Krifka et al. (1995: 14-16).

(72)a. The madrigal is popular.

b. Madrigals are popular.

c. ??A madrigal is popular.

(Krifka et al. 1995: 13)

Although characterizing sentences express general rules or a property that members of a certain class have, they allow for exceptions. Let's look at the following sentence that was used in Experiment 1 (73), combined with its story (74), to illustrate this property:

(73) Nessa escola *e* não pode escovar os dentes depois de comer. *BP*
In.this school not can:3SG brush:INF the teeth after of eat:INF
'In this school one cannot brush their teeth after eating.'

(74) Another weird thing happened in the school. The students were visited by a dentist. The dentist told the students that they should not brush their teeth after eating, *except for Joaquim,* who had to brush his teeth because they were too big. After lunch, the students almost forgot that they should not brush their teeth, but at the end they remembered the rule. Joaquim had to brush his teeth, because they were too big. He almost forgot about the recommendation given by the dentist, but he ended up remembering it. So Joaquim went to the restroom to brush his teeth after lunch. Bruno went to the play yard after lunch without brushing his teeth. Mariana was reading a book and she also didn't brush her

teeth. Lucas also didn't brush his teeth after lunch and went out to play with Bruno. Do you see how the students followed the rule?

The story in (74) shows a "weird" rule in the school: the students are not allowed to brush their teeth after lunch. The sentence in (73) states that it is a property of those who are subject to the rules of the school that they cannot brush their teeth. But, as we can see in the story in (74), even if there is an exception to this rule, namely, the student Joaquim, the rule still holds true in the school. The expression of a rule or a general fact with the allowance for exceptions is a property shared by characterizing sentences, as illustrated by the following example in Krifta et al. (1995)[20]:

(75) John smokes a cigar after dinner.

The sentence in (75) states that John has the habit to smoke a cigar after dinner: it reports the regularity of an episode. This sentence can be true even if from time to time John does not smoke a cigar after dinner, that is, even if there are exceptions to this generalization. However, if John *usually* does not smoke a cigar, the sentence will be false. Therefore, any approach to characterizing sentences should be able to capture the *quasi-universality* of these sentences, that

[20] It is possible, though, to have characterizing sentences that do not allow for exceptions. In (1), the proposition conveyed is true for *all* unmarried male adults (Papafragou 2016: 9). The characterizing sentence in (2), which expresses a mathematical concept, also does not allow for exceptions (Papafragou 2016: 28):
(1) A bachelor is unmarried.
(2) Two dots are joined by a single line.

is, the states of affairs denoted by the generic proposition needs to hold in the majority of the cases (Papafragou 1996).

There are many approaches to generic sentences. For instance, approaches that argue that genericity can be treated in terms of universal quantification, prototypes, stereotypes etc.[21] I will adopt here a modality-based approach to generics (Dahl 1975; Heim 1982), in which the generic operator is a modal operator. For us, the most obvious reason to adopt this approach is that modals, particularly deontic modals, seem to be acting as an overt expression of genericity in impersonal constructions in BP, suggesting that there is a link between genericity and modality.

In adopting a modal approach, we are assuming that characterizing sentences do not directly describe the "reality", but alternative state of affairs (i.e., possible worlds). For example, as was mentioned previously, the impersonal sentences used in our experiments describe "ideal" worlds in which a set of rules hold in a school.

There are three parameters that determine the meaning of the modal operator (Kratzer 1981) and two of them are relevant for us: modal relation and modal base. The modal relation includes the notions of possibility and necessity. The modal base (or conversation background) can be epistemic when a modality is understood as "in view of what is known", deontic when a modality is understood as "in view of what is commanded", bulletic, when it is understood as "in view of ones's wishes" etc. According to Heim (1982), the generic operator that is part of the logical form of characterizing sentences is a modal of necessity. In (76), "is a dog" and "has four legs" are predicates with a variable x bound by the silent operator GEN.

[21] For a overview of these approaches and their problems, see Krifka et al. (1995) and Papafragou (1996).

(76) Dogs have four legs.

GENx (x is a dog; x has four legs)

(Papafragou 1996:16)

The sentence in (76) conveys that in every world close to normal, all dogs have four legs. Exceptions will not make the statement in (76) false, since the sentence only conveys that it is expected that dogs have four legs.

Consider the following example below (77) that can be more easily constructed as a parallel to the sentences used in my experiments:

(77) A Christian is forgiving.

GENx (x is a Christian; x is forgiving)

(Papafragou 1996:17)

The sentence in (77) does not make a statement about the actual world, but about an ideal world in which every Christian is forgiving. Here, the operator GEN qualifies over all individuals in all worlds where all moral and religious obligations hold. The exceptions to the generalization in (77) would be the actual non-forgiving Christians that are not part of this ideal world (Papafragou, 1996: 17). Similarly to the previous cases, the allowance for exceptions does not make the sentence in (77) false. Now, consider the sentence in (73), repeated here in (78):

(78) Nessa escola não *e* pode escovar os dentes depois de comer. *BP*

In.this school not can:3SG brush:INF the teeth after of eat:INF

'In this school one cannot brush their teeth after eating.'

GENx (x is in this school; ¬ x brush their teeth after eating)

The modal base for the sentence in (78) is deontic, reflecting the rules of a particular school, and it contains a rule about brushing teeth. The sentence does not make a statement about concrete states of affairs in the actual world, but about rules holding in the actual world in which in the most normal circumstance no one in the school can brush their teeth after eating. The exception to the generalization in (78) would be the student Joaquim that for a particular reason is allowed to "break" this rule. Joaquim does not make the rule in (78) false, though, since (78) conveys not particular instances of the actual world, but a generalization about an ideal world in which the rule holds.

In the sentence in (78), the modal base is overt. I propose that in BP, for an impersonal sentence to have the generic reading, some overt marker of genericity is needed, such as a modal. There are countless examples of impersonal in BP lacking a modal described as acceptable by the literature, but these sentences, when accepted without any expression that conveys genericity have the existential reading. To give one of these examples from the literature, in Pilati, Neves and Salles (2005:122), the following sentence is judged as acceptable with the generic (inclusive) reading of the pronoun (the authors translate the pronoun as 'one'), but Carvalho (2018; 2019) argues that the interpretation of this sentence is not accurate: instead of *one*, which conveys the generic reading, the best translation for the null pronoun would be *someone*, which conveys the

existential reading. This existential reading is possible in BP if the impersonal sentence does not describe any kind of rule or generalization:

(79) Aqui *e* faz conserto de roupas. *BP*
Here make:3SG repair of clothes
≠ 'Here one fixes clothes.'
= 'Here someone fixes clothes.'

Consider now the sentence in (80), from Carvalho (2018: 83). The sentence lacks an overt modal, but there is an overt marker of genericity: the expression "it works in this way", which is followed by a description of a certain rule. The expression "it works in this way" is an overt marker of genericity in the sense that its presence helps us establish that the statement that follows it has a deontic modal base: it describes a world in which a rule holds (it also describes the consequences someone would suffer when disobeying the rule):

(80) Funciona assim: não comeu toda a comida, fica de castigo.
It.works like.this: not eat:PAST.SG all the food stay:3SG of punishment
'If one does not eat all the food, one is grounded.'

Although the sentence in (80) lacks an overt modal, it can be said to have a covert modal base that is deontic, as in the sentence in (78). The sentence states a rule that holds in a location that is unspecified, but supposedly understood by the addressee (e.g., in this house): one should

eat all the food. The fact that the sentence in (80) explicitly says that someone should endure some sort of punishment if the rule is disobeyed does not mean that the rule disallows exceptions. It merely states that in normal circumstances someone should eat all the food and that again, in normal circumstances, anyone that does not obey this rule is grounded. We could imagine a situation in which some individual is exempt from this rule due to special circumstances. Let's take Joaquim as our usual case of rule exemption: Joaquim does not need to eat all the food because he has lactose intolerance and "all the food" includes cheese, or Joaquim has diabetes and "all the food" includes rice with raisins, which he should cautiously avoid, or Joaquim is healthy, but he cannot eat all the food because his dog ate all the food he was supposed to eat. All sorts of situations can be imagined to illustrate that the rule in (80) holds even with exceptions. Due to the fact that impersonals like (80) have an interpretation that is very similar to the one that was spelled out for (78), it is natural to assume that both sentences introduce a GEN-operator. On the other hand, the sentence in (79) does not introduce a GEN-operator, but an existential operator instead.

Sentences like (78) with an overt modal and a null subject have *only* the generic reading in BP. If we observe the story in (74) that pairs with sentence (78), the reading in which the reference of the null subject could be constructed as the specific students from the school would be pragmatically felicitous. However, semantically, the test sentences used in the experiment can be classified as characterizing sentences which have a generic reading, not an existential one.

To make my proposal about the role of the modal in impersonal constructions in BP clear, let's consider the sentences in (81), given a lawlike background (there is a rule that allows the

students to bring toys). The sentences have a generic reading, considering the background and the fact that an exemption to the rule is potentially allowed:

(81) a. *Nessa escola traz brinquedo. *BP*

In.this school bring:3SG toy

'In this school one brings toys.'

GENx (x is in this school; x brings toys)

b. Nessa escola pode trazer brinquedo.

In.this school can:3SG bring:INF toy

'In this school one can bring toys.'

GENx (x is in this school; x brings toys)

Sentence (81b) has an overt modal, while sentence (81a) does not. The existential interpretation ("the group of students in the school") of the sentences in (81) is excluded although we could recover the reference of the pronoun from the context, as both (81a) and (81b) express a rule about the school, leading to the generic reading. However, sentence (81a) is predicted to be unacceptable, since BP requires an overt marker of genericity. An overt modal would be a marker of genericity, in the same way as adverbs like *always* or *typically*. The clitic *se* would be another overt marker of genericity. The sentence in (82) with *se* cannot be constructed as existential, only as generic:

(82) Nessa escola *se* traz brinquedo. *BP*

In.this school SE bring:INF toy

'In this school one brings toy'

≠ ∃x (x in in this school; x brings toys)

= GENx (x is in this school; x brings toys)

The generic reading of *se* is associated with the present tense and imperfective aspect. This same clitic can have the existential reading when associated with the past tense and perfective aspect (Carvalho 2018). Therefore, it is not only the presence of *se* which makes the sentence in (82) generic, but combination of *se* with the present tense and imperfective aspect.

In this section, we have seen that when no overt marker of genericity is present in an impersonal structure in BP and the sentence is not a lawlike statement, the impersonal structure has the existential reading for speakers. When the impersonal structure is a lawlike statement, but it doesn't have an overt marker of genericity, the sentence is judged at least as marginal (the survey and Experiment 2 to be discussed in Chapter 3 and 5, respectively, support this claim).

6 Conclusion

We saw that impersonal structures in BP can have a null subject and the overt *se* pronoun. Other overt pronouns are possible in impersonal structures, although they are not the focus of this thesis. These pronouns are: 'você' (*you*), when the impersonal structure has an inclusive generic reading, and 'eles' (*they*) when impersonals have a generic exclusive reading.

As the test sentences used in the experiments reported in Chapters 4 and 5 are lawlike statements, they have a generic reading. When a null impersonal has a generic reading, the null subject will be a φP pronoun.

Holmberg (2005, 2010a,b) proposes that in partial null-subject languages like BP, the EPP is checked by an element other than the goal of T's probing (e.g., by an adverbial). As discussed, there is no evidence that an EPP feature needs to be checked in generic impersonal constructions. Very often, generic constructions appear with a locative, but for reasons that are independent of the EPP: locatives appear to fulfill a semantic requirement to have the place where the rule-like statement holds specified whenever it is necessary.

We saw that in BP, an overt marker of genericity is required (or at least preferred) in generic impersonals. These overt markers of genericity can assume a variety of forms, but the ones present in my experiments were an overt modal and the *se* impersonal pronoun. In the absence of a genericity marker, the sentence has an existential reading, when there is no lawlike background forcing the generic reading of the sentence. It is a question that needs to be addressed in the future why BP overtly marks genericity in impersonal structures - why is a lawlike background not enough in generic sentences?

Chapter 3: Survey on Impersonal Structures in BP

0 Introduction

In this chapter, I report the results of an online survey conducted with adult native speakers of BP. Respondents rated auditorily-presented sentences given in context on a 5-point scale (to be described in more detail below). The main purpose of the survey was to see if the hypotheses raised in the first chapter were supported, namely:

(i) In generic impersonal constructions, the EPP needs to be checked by an element other than the null pronoun which values T's φ-features.

(ii) The generic reading of impersonal sentences requires an overt marker of genericity.

The hypothesis in (i) predicts that if there is in fact an EPP feature that needs to checked, speakers should not accept sentences such as (1a), which have the generic reading but no element to check the EPP.

(1)a. Não pode largar a escola.
Not can:3SG drop:INF the school
'One should not drop out of school.'

b. Nessa escola *e* não pode comer banana.
In.this school not can eat:INF banana
'In this school one cannot eat bananas.'

The hypothesis in (ii) says that the generic reading is possible only in sentences with overt markers of genericity (such *se* and a deontic modal). This hypothesis predicts that sentence (1b) should be acceptable, since the sentence has a deontic modal which functions as an overt marker of genericity.

To assess the hypothesis in (ii), generic structures with a deontic modal or generic *se* were compared with generic structures without these elements. The prediction was that generic structures without an overt marker of genericity would not be acceptable, under (ii).

This chapter is organized in the following way: in section 1, I explain the methodology and the materials adopted in the survey. Section 2 describes the participants in the study. In section 3, I report the results of the survey (descriptive statistics). Section 4 separately discusses the results of some of the sentences in the survey which had the existential reading instead of the generic one. Section 5 discusses whether our data support the hypotheses in (i) and (ii) (inferential statistics). Section 6 concludes the chapter.

1 Methodology and Material

The survey consisted of an Acceptability Judgment Test, conducted online. The survey was hosted by *Qualtrics*, a web-based platform that allows researchers to create surveys and generate reports. *Qualtrics* supports many languages, including Portuguese, in such a way that I was able to run this survey using only the participants' native language.

Table 1 shows the test sentences that were used in the survey together with their code representing the type of sentence (column 1) and the code explanation (column 2) (e.g., ADV+D (code), adverb + deontic modal (code explanation)). Additionally, five fillers were presented

randomly in the survey (see Appendix A). One of the fillers was categorically ungrammatical, two of them were completely grammatical and two of them admitted gradation in the judgement. In order to keep the survey short and increase the rate of completion, no more than 21 sentences were used[22]. The average time to complete the survey was 13 minutes. Notice that there were two sentences (in the conditions ADV and NoADV) in the survey that were compatible with the existential reading of the null subject: they will be analyzed separately (see section 4).[23]

Table 1: sentences used in the survey[24]

Code	Code Explanation	Sentences
ADV	Adverb	Nessa loja (se) vende saia. In.this store sell:3SG skirt *Null subject:* 'In this store someone sells skirts.' *Se:* 'In this store one sells skirt.'
		Nessa universidade (se) estuda na biblioteca. In.this university study:3SG in.the library 'In this university one studies in the library.'
NEG	Negation	Não (se) faz bolo de cenoura desse jeito. Not make:3SG cake of carrot in.this way 'One doesn't make carrot cake in this way.'
		Não (se) vende mais geladinho. Not sell:3SG more ice.pop 'One doesn't sell ice pops anymore.'
ADV+NEG	Adverb + Negation	No Brasil não (se) colhe cogumelo em bosques. In.the Brazil not pick:3SG mushroom in forests 'In Brazil one does not pick mushrooms in forests.'

[22] Two other sentences were included, but they are not reported here. The total number of sentences analyzed in this chapter is 19.

[23] At the time the survey was designed I was unaware of Carvalho's (2018; 2019) observation that some impersonal structures can have the existential reading in BP.

[24] Where it is indicated that the existential reading of the pronoun is possible (by using the pronoun *someone* in the translation), it only happens in null impersonals, not in impersonals with *se*. As seen in the previous chapter (Chapter 2, section 5), the clitic *se* can only have a generic reading when associated with the present tense and imperfective aspect.

Code	Code Explanation	Sentences
		Nos Estados Unidos não (se) assiste novela. In.the States United not watch:3SG soap.opera 'In the US one does not watch soap operas.'
ADV+D	Adverb + Deontic Modal	Nessa escola (se) tem que comer verdura na hora do almoço. In.this school have:3SG that eat vegetables at.the time of lunch 'In this school one has to eat vegetables at lunch time.'
		Nessa escola (se) tem que entregar lição de casa. In.this school have:3SG that turn.in:INF lesson of home 'In this school one has to turn in homework assignments.'
NEG+D	Negation + Deontic Modal	Não (se) pode largar a escola. Not can:3SG leave:INF the school 'One cannot drop out of school.'
		Não (se) pode andar de bicicleta na calçada. Not can walk:INF of bicycle in.the sidewalk 'One cannot ride a bicycle on the sidewalk.'
ADV+NEG+D	Adverb + Negation + Deontic Modal	Nessa escola não (se) pode trazer animal de estimação. In.this school not can:3SG bring:INF pet 'In this school one cannot bring pets.'
		Em Samoa não (se) pode esquecer o aniversário da esposa. In Samoa not can:3SG forget:INF the birthday of.the wife 'In Samoa one cannot forget one's wife's birthday.'
NoADV	No Adverb	(Se) vende doce. sell:3SG dessert *Null subject:* 'Someone sells desserts.' *Se:* 'In this store one sells desserts.'
		(Se) toma banho todo dia. take:3SG shower every day 'One shower every day.'

It is important to mention that all sentences were presented with a context initially given (see Appendix A), as the example below shows:

(2) The world is full of husbands with poor memory, but Samoa, a country in Oceania, is trying to change it! In Samoa a husband is forbidden to forget his wife's birthday. If he does, he is brought to justice and has to indemnify his wife for his actions.

(3) Em Samoa não (se) pode esquecer o aniversário da esposa.
In Samoa not can:3SG forget:INF the birthday of.the wife
'In Samoa one cannot forget one's wife's birthday.'

The participants listened to an audio with the context and the sentence in BP (no transcription was provided). The participants were informed in the initial screen that they had to judge the form of the sentence and not its content, providing judgements on a scale from 1 (very bad) to 5 (perfect). Half of the participants listened to the test sentences with *se* and the other half listened to identical test sentences, but without *se.*

For each one of the structures presented in Table 1, I summarize below which judgements were predicted under the hypothesis (i) and (ii) for the sentences *with se* and *without se*:

ADV: under the hypothesis in (i), both the sentences *with* se and *without se* should be judged as grammatical, since the adverbial can check the EPP. However, under the hypothesis in (ii), we expected that sentences *with se* would be judged as better than sentences *without se*, since I predict that the generic reading of impersonal sentences in BP requires an overt marker of genericity.

NEG: sentences in this condition should be judged as ungrammatical, since there is no element to check the EPP (hypothesis (i)).

ADV+NEG: under the hypothesis in (i), sentences in this condition should be judged as grammatical, since the adverbial can check the EPP. Under the hypothesis in (ii), sentences *with se* would be judged as better than sentences *without se* for exactly the same reason as in ADV: BP-speaking adults prefer generic structures with an overt marker of genericity than without it.

As *se* is an overt marker of genericity, ADV+NEG is expected to be better with *se* than without it.

ADV+NEG+D: under the hypothesis in (i), sentences in this condition should be judged as grammatical, as the fronted adverbial can check the EPP. Under the hypothesis in (ii), sentences *without se* in this condition should be judged as equally good as sentences *with se* if nothing in the language prevents double markers of genericity (*se* and the modal).

ADV+D: under the hypothesis in (i), sentences in this condition should be judged as grammatical, since the adverbial can check the EPP. Under the hypothesis in (ii), sentences *without se* in this condition should be judged as equally good as sentences *with se* for the same reason as in ADV+NEG+D.

NEG+D: sentences in this condition should be judged as ungrammatical, since there is no element to check the EPP (hypothesis (i)).

NoADV: sentences in this condition should be judged as ungrammatical, since there is no element to check the EPP (hypothesis (i)).

2 Participants

A total of 22 participants had their answers analyzed. I excluded 3 participants that demonstrated lack of seriousness in the way they filled in the survey (two participants were excluded for judging all sentences, including fillers, as either very bad or perfect, and one participant was excluded because they took about 1 minute to complete the survey, which shows that they did not listen to the audios before providing a judgment). 11 of the participants that were included in the analysis listened to sentences *with se* (Test 1) and the other 11 participants

listened to the sentences *without se* (Test 2). The participants included in the analysis responded to the fillers as expected.

The participants were from many regions of Brazil, although the distribution is clearly unbalanced, with a predominance of participants from the Southeast region of Brazil (both for Test 1 and 2), as can be seen in the map below[25] [26]. As the number of participants per region is small and unbalanced, I will not report the results by region.

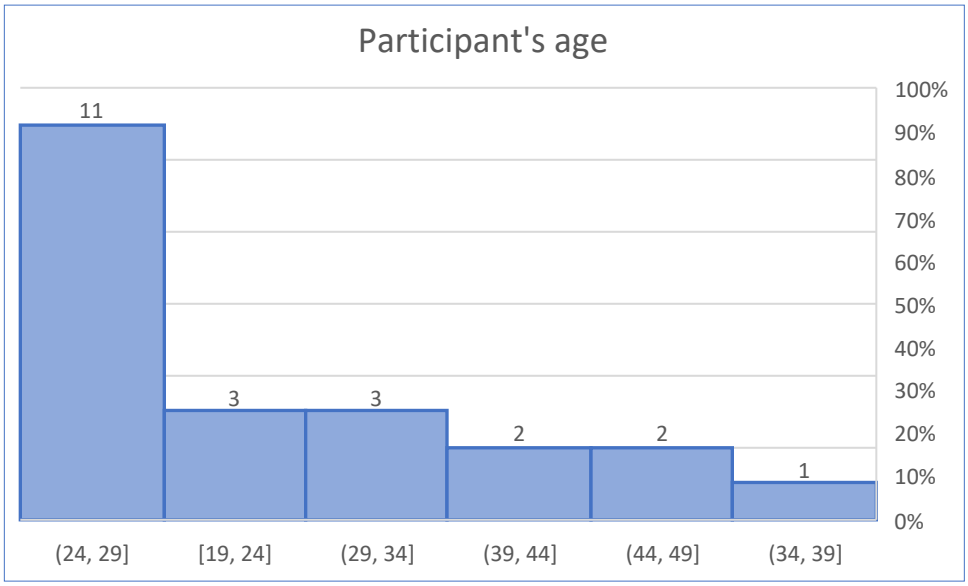
Graph 1: Participant's geographic distribution

[25] The percentage refers to the proportion of participants in each region, not in particular states. For example, even if the number 54.55% is in the state of São Paulo on the map, we are counting the proportion of participants in the whole Southeast region represented in orange.

[26] Minas Gerais is a state in the Southeast region, represented in pink in our map. The initial idea was to represent Minas Gerais separately, due to the particularity of their dialect that does not align with the rest of the Southeast region. This dialect is known to be innovative as regards null *vs* overt subjects in particular, and its loss of *se* that has been reported so often in the literature (see for example, Galves 1986 and Lemle 1985). However, only one participant out of 22 was from Minas Gerais. Therefore, it is impossible to extract any generalization about this dialect.

Participants were between the ages of 19 to 46. The mean age was 31.73 for those that took Test 1 (sentences *without se*) and 29.82 for those that took Test 2 (sentences *with se*). There is no significant difference in the age of participants that received Test 1 (*without se*) and Test 2 (*with se*): U = 57.5, p = 0.843 (2-tailed) (Mann-Whitney test). The histogram below shows the distribution of the participant's ages. As we can see, most of the participants were between 24- and 29-years-old (11 out of 22).

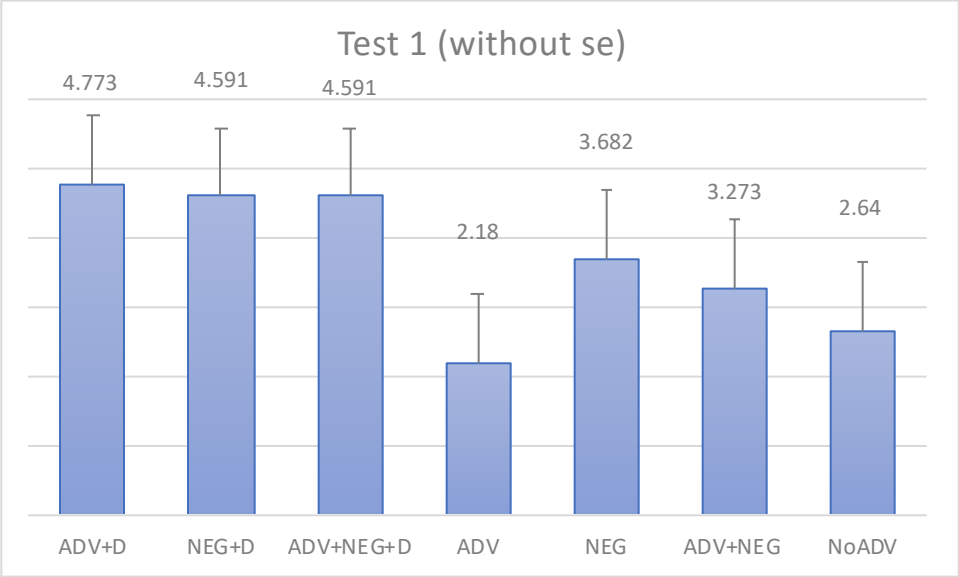
Graph 2: Histogram of age of participants



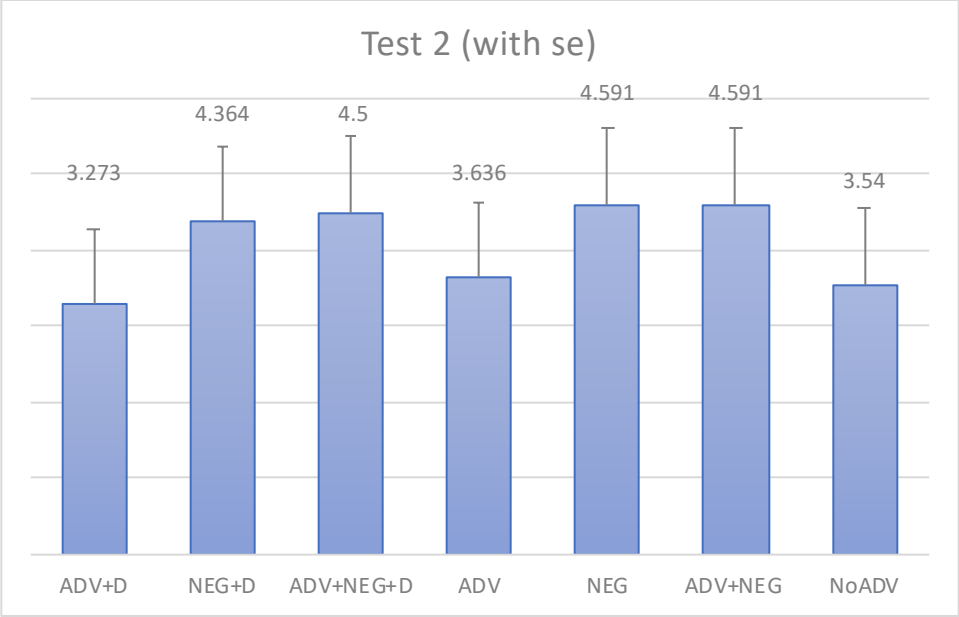
3 Descriptive Statistics

The bar-graphs below show the structures tested and their respective means on a 1 (very bad) to 5 (perfect) scale. The bar-graph excludes the sentences in the condition ADV and NoADV which had a context favoring the existential reading. These sentences are discussed separately in section 4.

Graph 3: Average for each structure: sentences without se (Test 1)[27]



Graph 4: Average for each structure: sentences with se (Test 2)



[27] Error bars represent standard deviation (SD).

The following table shows the means and standard deviations for the filler items. As expected, the sentence which was categorically ungrammatical received a very low score (≈ 1), while the two grammatical sentences received high scores (≈ 4-5). The ones that admitted gradation were judged as in between (≈ 3-4). The fillers were exactly the same for both Test 1 (*without se*) and Test 2 (*with se*)

Table 2: Averages and standard deviation for filler items

Grammaticality	**Test 1 (without *se*)**		**Test 2 (with *se*)**	
Average (AV) / (SD)	AV	SD	AV	SD
?	4.09	0.67	3.55	1.30
?	4.18	1.03	3.09	1.38
✓	5.00	0	4.45	0.89
✓	4.64	0.88	4.91	0.29
✗	1.18	0.57	1.45	1.16

I will first discuss the results for Test 1 (*without se*). As seen in Graph 3, sentences with an overt deontic modal and a null subject were judged as better than sentences without the modal. All the stories or contexts that preceded the test sentences had a lawlike background or described a tradition, or common behavior of people in a certain society (e.g., In Brazil, people don't pick mushroom in forests). These were sentences that had necessarily a generic reading, as can be seen in the example in (4):

(4)a. Brazilian soap operas are famous around the world, but this popularity did not reach every country. In 1995, the Yugoslav War stopped for 1 week so that the country could watch the last episodes of 'Slave Isaura'! But, as said before, it is not everywhere that people watch soap operas.

b. Nos Estados Unidos *e* não assiste novela.
In.the States United not watch:3SG soap.opera
'In the United States **one** doesn't watch soap operas.'

It would be reasonable to think that speakers gave a lower value in the 1 to 5 scale to the sentences in the conditions ADV, ADV+NEG and NEG compared to the sentences in the conditions ADV+D, ADV+NEG+D and NEG+D because generic sentences with an overt marker of genericity are at least preferred by speakers, but notice that they are not completely unacceptable. As ADV, ADV+NEG and NEG sentences lack an overt modal, the generic reading, in general, was not considered "perfect" by the participants. In this way, we could explain the difference seen in the results: sentences with a deontic modal received a high value in the judgement scale because they had a modal overtly marking the genericity of the sentences; on the other hand, sentences without a deontic modal received a lower value in the judgment scale because speakers needed an overt expression of genericity in these sentences.

The fact that the average score for the condition NEG+D was high (4.591) suggests that the EPP does not hold. There is no element in sentences in this condition that could check the EPP. Notice that NEG+D received the same score as ADV+NEG+D (4.591), a sentence in which

the EPP could be checked by the adverbial, as suggested in Holmberg (2005; 2010a,b), Rodrigues (2004); Pilati, Nave and Salles (2007); Holmberg et al. (2009).

Regarding the results for Test 2 (*with se*) (Graph 4), it is important to remember that the context that preceded these sentences was the same as the one that preceded sentences *without se*. The test sentences were also identical, the only difference being that these sentences had *se*, as we can see by the contrast between sentences (5) (Test 1) and (6) (Test 2):

(5) Não vende mais geladinho.
Not sell:3SG more freeze.pop
'They don't sell freeze pops anymore.'

(6) Não **se** vende mais geladinho.
Not SE sell:3SG more freeze.pop
'They don't sell freeze pops anymore.'

Considering hypothesis (i), we were expecting that the condition NEG (with *se* as an overt marker of genericity) in Test 2 would receive a low average score, in the same way as NEG+D (with the modal as an overt marker of genericity) in Test 1. In fact, the average score for the condition NEG (with *se*) was exactly the same for NEG+D (without *se*), as expected if the hypothesis that the EPP holds is not supported.

As is shown in Graph 4, the condition ADV+D in Test 2 received a lower average score compared to the other conditions with a modal: 3.27. We could not explain the low score

obtained for the condition ADV+D by saying that the modal cannot co-occur with *se*, because otherwise we would expect a low score for the conditions NEG+D and ADV+NEG+D, which did not happen: these two conditions received the near perfect scores of 4.36 and 4.5, respectively. Therefore, *se* can co-occur with the modal. Why did the condition ADV+D receive a low score then?

The condition ADV+D was the only one in which the modal *ter que*[28] ('have to') was used. All the other conditions with a modal (NEG+D and ADV+NEG+D) had the negated modal pode ('can'): *não pode* ('cannot'). It may be the case that there is some property in *tem que* ('have to') that disfavors its co-occurrence with *se*.[29]

The modal *tem que* ('have to') is not semantically equivalent to *não pode* ('cannot'), the other modal used in the survey, although both of them can be understood as necessity modals.[30] In the example in (7), *tem que* describes that in the school one has the *obligation* to bring pets. In the example in (8), *não pode* describes that in the school one is *not allowed* to bring pets:

(7) Nessa escola *e* tem:3SG que trazer animais de estimação.
In.this school *e* have that bring:INF pets
'In this school one has to bring pets.'

[28] *Ter que* is the infinitive form. The form used in the survey was *tem que* (present tense, third person).

[29] *Tem que* was also used in half of the test sentences in Experiment 1 (Chapter 4). But all the sentences in Experiment 1 had null subjects (they were sentences without*se*), causing no problem in the wellformedness of the sentences.

[30] We are comparing *tem que* to the negated version of *pode. Pode,* without a negator, is very often associated with an epistemic reading (equivalent to *may*), while *não pode* has a deontic reading. I discuss the difference between *pode* and *não pode* in more detail in Chapter 4.

(8) Nessa escola *e* não pode trazer animais de estimação.

In.this school not can:3SG bring:INF pets

'In this school one cannot bring pets.'

While I am pointing out a contrast between the two modals that could explain the difference found in the results of the survey (Test 2), the precise reason why a modal such as *tem que* cannot co-occur with *se* needs to be further investigated. Also, it would be necessary to have more solid evidence that *se* is in fact disfavored with *tem que.* As an online search for academic texts shows, *tem que* can co-occur with *se* at least in the formal written register:

(9) Os defensores do JIT apontam que estoques altos se faz necessário

The defenders of.the JIT point.out that stocks high SE make:3SG necessary

quando **se tem que** esconder problemas maiores na empresa.

when SE have that hide:INF problems bigger in.the company

'The defenders of the JIT point out that high stock prices are necessary when one has to hide bigger problems in the company.'[31]

(10) A formação de iniciantes na arqueologia nesse ambiente pode ser, portanto,

The training of beginners in.the archeology in.this environment might be therefore

bastante inadequada,pois o que **se tem que** aprender é a produzir ciência […]

quite inadequate because the what SE have that learn:INF is to produce science

[31] https://www.unimep.br/phpg/bibdig/pdfs/docs/11102013_144041_nataly.pdf

‘Therefore, in this environment, the training for beginners in archeology might be quite inadequate because what one has to learn is to produce science […].’[32]

In section 5, I will analyze the results more closely, in the attempt to investigate whether the data support the hypotheses formulated at the beginning of this chapter. But, before that, I will discuss the results for the sentences with the existential reading.

4 Impersonals with Existential Reading

One of the sentences in the condition ADV and another sentence in the condition NoADV admit the existential reading or even the definite one if the reference of the pronoun would be judged as “Joaquim’s mother” (11) and “Mariana’s father” (12): this could be possible if the owners of the store are the ones who sell the products. This reading, though, should be considered degraded, as in BP the definite reading of the third person pronoun in matrix clauses is usually only possible in subjectless replies or when the reference of the null subject can be recovered by the physical presence of the referent. The sentences could be analyzed as existential if some unspecified person sells the products in the stores. In this case, ADV should be judged as acceptable and NoADV not, as existential sentences require an obligatory event argument.

(11)a. Joaquim’s mother has a store called ‘Just skirt.’

b. Nessa loja vende saia.
In.this store sell:3SG skirt

[32] FUNARI, P. P. (2000). Como se tornar arqueólogo no Brasil. *Revista USP*, (44), 74-85. https://doi.org/10.11606/issn.2316-9036.v0i44p74-85

'In this store someone sells skirts.'

(12)a. Mariana's father has a store called 'Just dessert.'

b. Vende doce.

Sell:3SG dessert

'Someone sells desserts.'

The sentence in (11) (ADV) (Test 1, *without se*) received the average score of 4.91 (SD = 0.29). Assuming that the definite reading is not possible in (11), the existential reading was well-accepted. The sentence in (12) (NoADV) (Test 1, *without se*) received the average score of 3.64 (SD = 1.30). The results of the Wilcoxon signed-rank test show that the difference is significant ($Z = -2.8031$, $p < 0.05$, 2-tailed).

Regarding the results for the same sentences accompanied by a context which favored the existential reading, ADV (13) and NoADV (14), in Test 2 (*without se*), they received the average score of 4.09 (SD = 0.94) and 3.09 (SD = 1.14), respectively. The results of the Wilcoxon signed-rank test show that the difference is not statistically significant ($Z = -1.5401$, $p > 0.05$, 2-tailed). Recall that in Chapter 2 it was pointed out that the impersonal clitic *se* is associated with the generic reading when the sentence is in the present tense and imperfective aspect. As the context favoring the existential reading is incongruent with the generic reading of the sentences by themselves, one would expect that these sentences would be judged as unacceptable. This is a puzzle that needs to be addressed in the future. Perhaps speakers judged the sentences without

taking into account the context. However, as there were only two sentences with the existential context, I will refrain from making any strong claim about the results.

(13)a. Joaquim's mother has a store called 'Just skirt.'

b. Nessa loja **se** vende saia.

In.this store sell:3SG skirt

'In this store one sells skirts.'

(14)a. Mariana's father has a store called 'Just dessert.'

b. **Se** vende doce.

sell:3SG dessert

'One sells desserts.'

As pointed out, the difference in the average scores for each one of the conditions was not statistically significant. However, according to my judgment *se* sounds better as an enclitic (e.g., vende-**se**) than as a proclitic (e.g, **se** vende) when it occurs in the beginning of a sentence (in other contexts, *se* has the tendency to appear as a proclitic in BP), which can explain why NoADV (14) received a lower average score than ADV (13).

5 Inferential Statistics

Until now I described the results from the sample I obtained. Now we want to make inferences about the population studied based on that sample. We should keep in mind, though,

that our sample is small: with a larger sample, the inferences would be more solid. The hypotheses formulated at the beginning of the chapter were:

(i) In generic impersonal constructions, the EPP needs to be checked by an element other than the null pronoun which values T's φ-features.

(ii) The generic reading of impersonal sentences requires an overt marker of genericity.

Using the Anderson-Darling test, we see that the data do not follow a normal distribution: AD = 4.03, $p < 0.005$ for Test 1; AD = 5.11, $p < 0.005$ for Test 2. As the data do not follow a normal distribution, a non-parametric test, the Friedman Test, similar to the parametric Repeated Measure ANOVA, was used to detect differences across conditions within Test 1 and within Test 2. That is, one separate Friedman Test was run for each test.

For Test 1 (*without se*), when the conditions ADV, ADV+D, ADV+NEG+D, NEG+D, ADV+NEG, NEG and NoADV were compared, a significant difference was detected (chi-square = 18.52 $p < 0.01$, 2-tailed). A post hoc analysis was conducted to see in which conditions the difference was found. The results are reported in the box below. Conditions with the same letter are not significantly different:

Box 1: Post hoc analysis: sentences without se (Test 1)

Post Hoc Analysis
Alpha: 0.05; DF Error: 60
t-Student: 2.00298
LSD: 16.37314

Condition	Median	Mean	Sum of ranks	Groups

ADV+D	5	4.77	67.0	a
ADV+NEG+D	5	4.59	61.0	a
NEG+D	5	4.50	60.0	a
NEG	5	4.36	42.0	b
ADV+NEG	4	3.64	36.0	b
NoADV	4	3.54	26.5	c
ADV	3	3.27	15.5	d

I will first discuss whether the results support the hypothesis in (ii). I will address hypothesis (i) after this discussion. As explained before, while the sentences in the conditions ADV+D, NEG+D and ADV+NEG+D had an overt marker of genericity (i.e., the deontic modal), the sentences in ADV, NEG, ADV+NEG and NoADV did not have a overt marker of genericity, only a background favoring the generic reading. In Box 1, we see that the conditions that received a higher score (*group a*) are the ones with a modal in contrast to the ones without a modal (*group b, c and d*). This seems to point out that participants at least prefer sentences with an overt marker of genericity as opposite to sentences without it when sentences are presented with a background favoring the generic reading. The results do not suggest that the overt marker of genericity is *required* in generic sentences, only that it is preferred, as sentences without a modal were not judged as completely ungrammatical.[33]

Notice that ADV and NoADV received a significant lower score than the other sentences without a modal (still, all sentences without a modal were judged as worse than sentences

[33] In Experiment 2, reported in Chapter 5, the condition ADV had a test sentence without a modal and with a lawlike background. Participants had to choose between a null subject (no overt marker of genericity) and a *se* pronoun (overt marker of genericity). They chose *se* 100% of the time. However, by the nature of the experiment, which asked for the participants' *preference*, there is no way to know if they would *exclude* sentences without the overt marker of genericity.

without it). Recall that there was only one generic sentence for each one of these conditions, while for the other conditions, there were two items. A balance in the number of sentences per condition could have rendered different results.

The condition NEG+D was important for testing hypothesis (i), namely, that the EPP should be satisfied by some element other than the generic null pronoun. The prediction is that sentences in this condition should be unacceptable, if there is an EPP feature that needs to be checked by elements such as locatives occurring in [Spec, TP]. In Box 1, we also see that the condition NEG+D was not significantly different from generic structures with an adverbial that could check the EPP. It is necessary to take into account, though, that there was only one condition in Test 1 that tested the role of the EPP. The sentences in the conditions NoADV and NEG did not have an element that could check the EPP, but they also did not have an overt marker of genericity, in such a way that the absence of an overt marker of genericity is enough to explain the low average score obtained for these two conditions.

For Test 2 (*with se*), a significant difference was detected when the conditions ADV, ADV+D, ADV+NEG+D, NEG+D, ADV+NEG, NEG and NoADV were compared (chi-square = 18.51, $p < 0.01$, 2-tailed). A post hoc analysis was run to detect which conditions differ from each other. The results of the post hoc test can be seen below in Box 2. Conditions that share the same letter are not significantly different:

Box 2: Post hoc analysis: sentences with se (Test 2)

Post Hoc Analysis
Alpha: 0.05; DF Error: 60
t-Student: 2.00298
LSD: 9.382228

Condition	Median	Mean	Sum of ranks	Groups
NEG	5	4.72	55.5	a
ADV+NEG	5	4.59	52.5	a
ADV+NEG+D	5	3.27	52.0	a
NEG+D	4	3.68	52.0	a
ADV	3	3.27	35.5	b
NoADV	4	3.54	31.0	b
ADV+D	2	2.18	29.5	b

In the previous chapter (Chapter 2) it was proposed that *se* and the deontic modal function as overt markers of genericity. It seems, though, that nothing prevents *se* from co-occurring with a modal, given the fact that the conditions ADV+NEG+D and NEG+D are not significantly different from the conditions with the highest score without a deontic modal (ADV+NEG and NEG).

Participants were reluctant in accepting *tem que* (have to) with *se,* since the condition ADV+D was judged as bad as the condition NoADV and ADV (in which there was only one sentence per condition): it was pointed out above that *tem que* (have to) has a different interpretation than *não pode* (cannot) which was used for the conditions ADV+NEG+D and NEG+D. That might be the reason why participants in general rejected *tem que* with *se*. This result warrants further investigation to examine whether *tem que* in fact cannot co-occur with *se* and which properties are responsible for that.

The conditions ADV and NoADV in Test 2 are likely to have received a low average score compared to the sentences for other conditions for the same reason as in Test 1: the

conditions ADV and NoADV had only one sentence each, while the other conditions had two sentences each. A balance in the number of sentences per condition could have rendered different results.

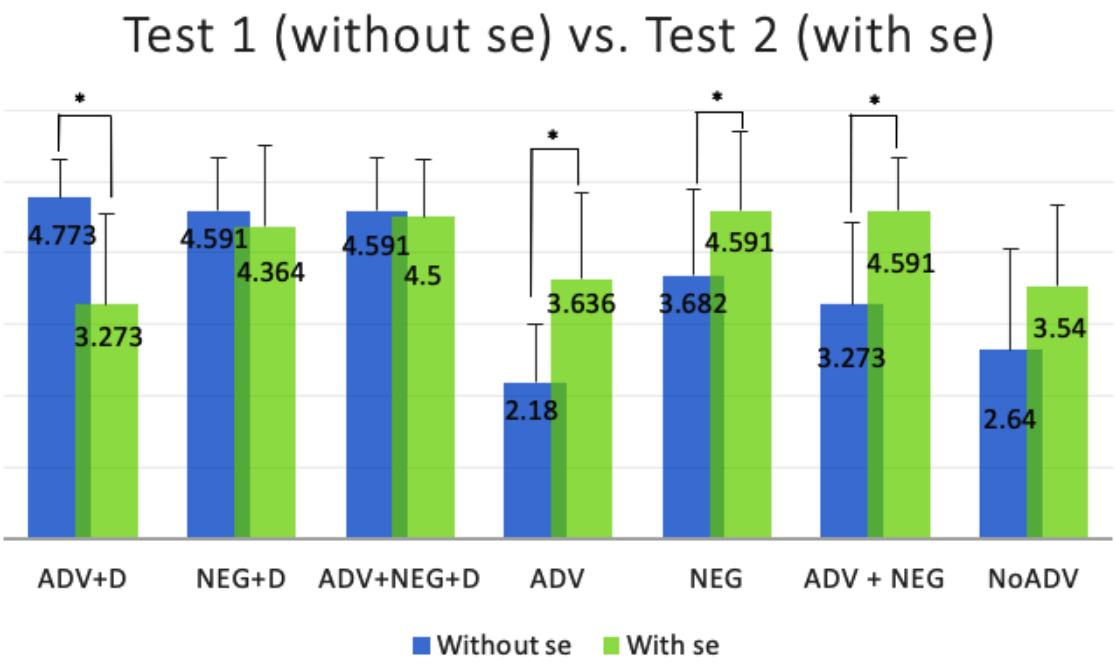
Notice that the score for the condition NEG and NEG+D were statistically identical to the conditions ADV+NEG, ADV+NEG+D and there is no adverb in the sentences for the conditions NEG and NEG+D that could check the EPP feature. The hypothesis (i) is not supported by the results of Test 2 and it is not supported by the results of Test 1 as well, as discussed above.

In Graph 5, we plotted the bar graphs for Test 1 (*without se*, in blue) and Test 2 (*with se,* in green) together. Examining Graph 5, a clearer pattern emerges and we are able to better analyze whether the results support the hypothesis (ii) or not (i.e., generic impersonal structures require an overt marker of genericity). Notice that all conditions without a modal, with the exception of NoADV, were judged as significantly better when *se* was included (Test 2) than when just a null subject was present (Test 1). These results suggest that the overt marker of genericity improves the availability of the generic reading. However, as sentences without an overt marker of genericity were not judged as completely ungrammatical, we can only say that BP-speakers *prefer* generic sentences with an overt marker of genericity.

In order to support the the hypothesis that sentences without an overt marker of genericity are not completely ungrammatical, I conducted the Friedman for Test 1 (*without se*) and Test 2 (*with se*) including now the ungrammatical filler. The prediction is that a significant difference should be found between the ungrammatical filler and all the other sentences, specially the ones without an overt marker of genericity. For Test 1 (*without se*), a significant difference was found between conditions (chi-square = 63.34, $p < 0.001$). The post hoc test detected a significant difference between the fillers and all the other conditions, with the the filler receiving the lowest average score (1.18).

Although all the sentences in Test 2 (*with se*) had an overt marker of genericity, the Friedman test including the ungrammatical filler was also conducted to verify the grammaticality of the test sentences. For Test 1 (*with se*), a significant difference was found between conditions (chi-square = 3.17, $p < 0.001$). The post hoc test detected a significant difference between the fillers and all the other conditions, with the the filler receiving the lowest average score (1.45).

Graph 5: Average for each structure: sentences without se and with se



Before proceeding in explaining the pattern seen in Graph 5, I describe how the statistical analysis comparing the two tests (Test 1 *vs* Test 2) was conducted. Pairwise Mann-Whitney tests were performed to compare each one of the conditions in Test 1 and Test 2. The Mann-Whitney test was chosen because the samples are independent (different participants took Test 1 and Test 2). As only one comparison was performed by condition, the tests were performed without a correction. The table below shows the values of U and the p-values for each one of the comparisons (e.g., ADV (1) and ADV (2) stands for ADV (Test 1) and ADV (Test 2), respectively).

Table 3: Values of U and p-values for Test 1 (*without se*) vs. Test 2 (*with se*)

Conditions	**Values of U**	**P-values (2-tailed)**
ADV (1) - ADV (2)	22	$p < 0.05$
NEG (1) - NEG (2)	24	$p < 0.05$
ADV+NEG (1) - ADV+NEG (2)	18.5	$p < 0.05$
ADV+NEG+D (1) - ADV+NEG+D (2)	55.55	$p = 0.76$
ADV+D (1) - ADV+D (2)	17	$p < 0.05$
NEG+D (1) - NEG+D (2)	59	$p = 0.94$
NoADV (1) - NoADV (2)	34	$p = 0.87$

I explain in detail the results for each one of the conditions below:

ADV: we expected that sentences *with se* would be judged as better than sentences *without se*, considering the hypothesis that the generic reading of impersonal sentences requires an overt marker of genericity in BP. As *se* is an overt marker of genericity, ADV is expected to be better *with se* than *without* it, as we would have no other overt marker of genericity in this kind of sentence. The test sentence *with se* rendered a higher score than the test sentences *without se* and the difference was statistically significant. The results do not support the hypothesis that the generic marker is *required*, but they suggest that speakers of BP *prefer* generic sentences with the overt marker of genericity.

Notice that the average score for this condition was low in both Test 1 and Test 2. It was pointed out that a factor that might have contributed to the low average score obtained for these conditions is that there was only one sentence for ADV (the other sentence was excluded from

the statistical analysis because they had an existential reading), while all the other conditions had two sentences each (with the exception of NoADV).

NEG: sentences *with se* in this condition should be judged as better than sentences *without se* for exactly the same reason as in ADV: adult speakers prefer generic sentences with an overt marker of genericity. As *se* is an overt marker of genericity, NEG is expected to be better *with se* than *without* it. For this condition, the test sentences *with se* were judged as significantly better than sentences *without se*, as predicted.

ADV+NEG: sentences *with se* in this condition would be judged as better than sentences *without se* for exactly the same reason as in ADV and NEG: adult speakers prefer generic structures with an overt marker of genericity than without it. As *se* is an overt marker of genericity, ADV+NEG is expected to be better with *se* than without it. For this condition, the test sentences *with se* were judged as significantly better than sentences *without se*, as predicted.

ADV+NEG+D: sentences *without se* in this condition should be judged as equally good as sentences *with se* if nothing in the language prevents double markers of genericity (*se* and the modal). It seems that there is nothing that prevents it, since the score was almost at ceiling in both Test 1 (*without se*, with no double marker of genericity) and Test 2 (*with se,* with a double marker of genericity). For this condition, no difference was found between the test sentences *with se* and *without se*, as predicted.

ADV+D: sentences *without se* in this condition should be judged as equally good as sentences *with se* for the same reason as in ADV+NEG+D. However, as explained before, the deontic modal *tem que* (‘have to’) was used in ADV+D, while the deontic modal used in all the other conditions was *não pode* (‘cannot’). It was suggested that speakers do not favor the co-

occurrence of *tem que* ('have to') with *se,* explaining why sentences *without se* were judged as better than sentences *with se.* This pattern clearly needs to be investigated further.

NEG+D: sentences *without se* in this condition should be judged as equally good as sentences *with se* if nothing in the language prevents double markers of genericity (*se* and the modal), as in ADV+NEG+D and ADV+D. For this condition, no difference was found between the test sentences *with se* and *without se*, as predicted.

NoADV: the results for this condition were very similar to the results for the condition ADV and this condition had the same problem as ADV (i.e., just one sentence for this condition). Overall, sentences *with se* were judged as better than sentences *without se*, but this difference was not statistically significant. A larger sample and more sentences in this conditions could have rendered significant results.

The average score for this condition was low in both Test 1 and Test 2. Two factors might be responsible for that: (i) there was only one sentence for NoADV and (ii) in the beginning of a sentence, *se* has the tendency to appear as an enclitic rather than a proclitic.

6 Conclusion

This chapter reported the results of an online survey conducted with adult native speakers of BP. The purpose of this chapter was to see if the following hypotheses were supported:

(i) In generic impersonal constructions, the EPP needs to be checked by an element other than the null pronoun which values T's φ-features.

(ii) The generic reading of impersonal sentences requires an overt marker of genericity.

Regarding (i), the results of the survey do not support the hypothesis that the EPP must be checked by elements such as adverbials in generic impersonal structures. The condition NEG+D (Test 1) and NEG (Test 2) were employed to check the role of the EPP: the prediction was that test sentences in these conditions would be unacceptable for the participants, because there is no element that can check the EPP in these structures. However, the average score for both of these conditions was almost at ceiling.

The hypothesis (ii) was shown to be too strong: the results only suggest that speakers *prefer* sentences with an overt marker of genericity, not that they *require* these markers. That speakers prefer generic sentences with an overt marker of genericity is supported by the following results: when comparing Test 1 (*without se*) with Test 2 (*with se*), we saw that in sentences without a modal (ADV, NEG and ADV+NEG, NoADV), the presence of an overt marker of genericity (*se*) in general improved the judgement of the sentences. Also, in Test 1 (*without se*), sentences with the deontic modal were judged as better than sentences without it.

Chapter 4: Experiment 1

0 Introduction

In this chapter I will present the first experiment I conducted with children acquiring Brazilian Portuguese as their native language, as well as with adults (control group). The experiment was designed to test if Brazilian children know that they are acquiring a partial null-subject language. My prediction was if children knew they were acquiring a partial null-subject language, they should reject the definite reading of the null subject in a sentence like (1a).

In partial null-subject languages such as BP, the null subject in the impersonal structure in (1a) can only be understood as generic: in the place specified by the fronted adverb 'in this school', *one* cannot brush one's teeth after eating. In European Portuguese, a consistent null-subject language, the sentence in (1a) only allows for the definite reading of the null pronoun (in which the null pronoun is understood as 'she or he'). In order to express the generic reading of the null subject, EP uses a *se* pronoun (1b):

(1)a Nessa escola *e* não pode escovar os dentes depois de comer. *BP*
In.this school. not can brush:INF the teeth after of eat:INF
'In this school one cannot brush one's teeth after eating.' [*in view of the laws*]

(1)b. Nessa escola não *se* pode escovar os dentes depois de comer. *EP*
In.this school not can brush:INF the teeth after of eat:INF
'In this school one cannot brush one's teeth after eating.'
[*in view of the laws*]

I designed the experiment described below in order to see if Brazilian children know that they are acquiring a partial null-subject language. The idea behind the experiment is that if children rejected the null subject of impersonal sentences without SE as definite and understood it as generic, we could assume that they knew they were acquiring a partial null-subject language.

Before describing the experiment, I would like to point out that radical pro-drop languages (languages without overt verbal agreement), such as Chinese and Thai, also allow only the generic reading of the null subject in sentences equivalent to (1a). In the following sentence in Chinese (2), the subject is null and it can only have the generic reading, the definite reading being impossible (see Holmberg and Phimsawat 2005 for Thai).[34]

(2) Zai zhe-ge xuexiao, *e* chi fan yihou *e* bu neng shua ya.
At this-CL school, eat meal after not can brush teeth
'In this school one cannot brush one's teeth after eating.'

As the prediction is that children will interpret the null subject in impersonal structures as generic in both radical pro-drop languages and partial null-subject languages, one could argue that the experiment to be reported will *only* be able to tell whether children acquiring BP know that they are *not* acquiring a consistent null-subject language like EP. However, it is known that BP-speaking children exhibit knowledge of agreement by the age of 3;0 (Magalhães 2006). Agreement is not present in radical pro-drop languages, but it is in partial null-subject languages.

[34] I am indebted to Shengyun Gu and Margaret Chui Yi for the Chinese data.

The agreement facts in conjunction with the interpretation of null subjects by Brazilian children can indicate whether they know they are acquiring a partial null-subject language or not.

1 Methodology and Material

The experiment consisted in a Truth-Value Judgement Task (TVJT) (Crain and McKee 1985). In this task, the child has to provide a binary judgment: true or false. In our case, in particular, children had to hear a sequence of narratives in BP about a school and judge, by the end of each story, if the character Elmo is saying something that happened in the story (true) or not (false). Power Point animations were used and the task took about 15 minutes to be completed. As already said, the objective of this experiment was to see if children understand that the null subject in an impersonal sentence has to be generic instead of definite. If they understand that BP is a partial null-subject language, the null subject would be judged invariantly as generic. Let's see below an example of one of the stories (see Appendix B for all the stories that were presented).

Slide 1 As you already noticed, this school is very weird. One of the rules is that just one of the students, Joaquim, has to eat dessert during lunch time instead of regular food.

Slide 2 Look at Bruno. He brought a salad to eat for lunch. Look, he is eating the salad now.

Slide 3 Mariana brought cheese balls to eat for lunch. Look, she is eating cheese balls.

Slide 4 Look at Lucas! He is eating pasta.

Slide 5 At first, Joaquim put a sandwich in his lunchbox, but then he remembered the rule that he had to bring dessert for lunch, so he brought a brigadeiro[35] instead. Look, he is eating a brigadeiro.

Slide 6 Now, Elmo is going to tell us a part of the story. Let's see whether he paid attention or not.

[35] A traditional Brazilian dessert, very popular among children.

Slide 7 Nessa escola e tem que comer doce na hora do almoço.

In.this school has that eat.INF dessert at time of.the lunch

'In this school one has to eat dessert for lunch.'

In the story, Joaquim is the possible definite subject. Notice that if the child understands the null subject as generic, they should say that the test sentence presented in the slide above is false, because in the school there was no general rule that children should eat dessert for lunch; this rule was just applied to a single character: Joaquim. On the other hand, if the child understands the null subject as definite, they would say that the sentence is true, because Joaquim, in fact, was the only one that had to eat dessert for lunch.

The experiment was composed of six test sentences (Table 1), two training items and two fillers, the last ones presented randomly between the test sentences. For half of the test sentences, the expected answer was 'true' because the generic reading was true in the context; for the other half the expected answer was 'false' because only the definite reading was possible (and this reading is ungrammatical for adults). All the test sentences were null impersonals with a fronted adverbial and a deontic modal, as we can see in Table 1. This structure was used because the survey discussed in Chapter 3 showed that null impersonals that followed this pattern were highly accepted by native speakers of BP. In this way, we could exclude confounding factors and better claim that the reason why children rejected the test sentences that were expected to be reject and accept those that were expected to be accepted was only because they rejected the definite reading of the null subject and accepted its generic reading.

Table 1: Test sentences, Experiment 1

CONDITION	TEST SENTENCE	EXPECTED ANSWER
Tem que ('have to')	Nessa escola *e* tem que trazer animal de estimação. In.this school has that bring:INF pets 'In this school one has to bring pets.'	T
Tem que	Nessa escola *e* tem que comer doce na hora do almoço. In.this school has that eat:INF dessert in.the time of.the lunch 'In this school one has to eat dessert for lunch.'	F

Tem que	Nessa escola *e* tem que brincar dentro da sala de aula. In.this school has that play:INF inside of.the room of class 'In this school one has to play in the classroom.'	F
Não pode ('cannot')	Nessa escola *e* não pode estudar de manhã. In.this school not can study:INF of morning 'In this school one cannot study in the morning.'	F
Não pode	Nessa escola *e* não pode escovar os dentes depois de comer. In.this school not can brush:INF the.PL teeth after of eat 'In this school one cannot brush the teeth after eating.'	T
Não pode	Nessa escola *e* não pode entregar a lição de casa. In.this school not can turn.in:INF the lesson of home 'In this school one cannot turn in the homework.'	T

Notice that the two deontic modals used in the test sentences were 'tem que' (have to) and *não pode* (cannot) . Some valid criticisms could be made regarding this choice. First of all, the modal with the negation (*não pode*) should have been avoided since the negation might introduce processing difficulties for the child. Also, it could be argued that as *não pode* is preceded by a negator and *tem que* does not have it, one cannot say that the test sentences are testing equally the same thing (i.e., the test sentences do not have two modals, but a modal (*tem que*) and a modal preceded by negation (*não pode*)).

There is a reason why negation was added to the second modal, though. The modal *tem que* very clearly denotes obligation and imposition (Comparini 2008), while *poder* by itself is very often associated with the epistemic reading, but when preceded by the negation, *poder* can be used as a "strong restrictor", a modal that imposes a prohibition (or a rule) and it is, therefore, clearly deontic (Comparini 2008). Another verb with deontic modality that could have been used was *dever* (*must*). Nevertheless, this modal can have an epistemic reading in some contexts (see (3) below, in which this modal expresses "possibility"):

(3)... e nessa hora... é que eu percebi que o que tinha... era uma coisa assim de de de::… uma mistura de sentimento que ***devia*** ter uma ligação com ciúme... porque eu nunca fui ciumenta... (Comparini (2008: 41)

'*…and at this moment… it was when I noticed what happened… was something like… a mix of feelings that* ***must*** *have an association with jealously… because I was never jealous…'*

It should be also pointed out that *dever* has different requirements when combined with the impersonal clitic *se*. In a diachronic study on pronominal placement, Nunes (2015: 163) observes that the clitic *se* survived as an enclitic in BP in some structures with *dever*, while it is a proclitic in structures with other modals and infinitive verbs. According to Nunes, that reflects the maintenance of a normative pattern. *Dever* not only has this different structural behavior compared to other modals regarding *se*; as a native speaker, I have the intuition that *dever* requires *se* in an impersonal structure more often than other modals. In sum, I chose to use *poder* with the negator because (i) it would be problematic to use *poder* by itself, without the negation, because the modal would probably be interpreted as epistemic by the participants; (ii) the other modal that could be used as deontic is ambiguous in some contexts and we couldn't be sure that it would be accepted in impersonals without *se*.

In order to have minimal-pairs, a follow-up study with the addition of *não tem que* (do not have to) could be conducted. It should be noticed, though, that it was not relevant for our purposes to contrast the items *tem que* and *não pode* (although similar results were obtained for both conditions), reducing the need to have minimal pairs. Another point is that, although *não tem que* and *não pode* are minimal pairs in the sense that both are modals preceded by a

negation, they have very different meanings: while *não pode* means that a person is not allowed to do something, *não tem que* means that a person doesn't have the obligation to do something.

As the results obtained for the condition with negation and without negation were statistically identical, it seems unlikely that the negator caused processing difficulties for the participants. The negation seems to assure that the modal *poder*, otherwise ambiguous, will have a deontic reading.[36]

The reader might think that a test including control sentences that tested the difference between epistemic modality and deontic modality should have been included. This would make sense to interpret the results of the first experiment if children performed at chance level. If so, we could hypothesize that the random results were due to the fact that children thought that the deontic modal from all the test sentences in Experiment 1 was actually epistemic, leading them to a state of confusion regarding the acceptability of the null impersonal. With that hypothesis we would be assuming that children know that in BP the "perfect" structure for a null impersonal is the one that contains an overt marker of genericity, a fact that was investigated in Experiment 2. However, as we will see, children's performance was adult-like in the first experiment: that makes the control sentences testing the difference between epistemic and deontic modality not necessary.

2 Participants

I interviewed forty children acquiring BP as their native language. They ranged in age from 4;0 to 7;10 (mean age = 5;8). The children were recruited in a variety of places: "Museu da

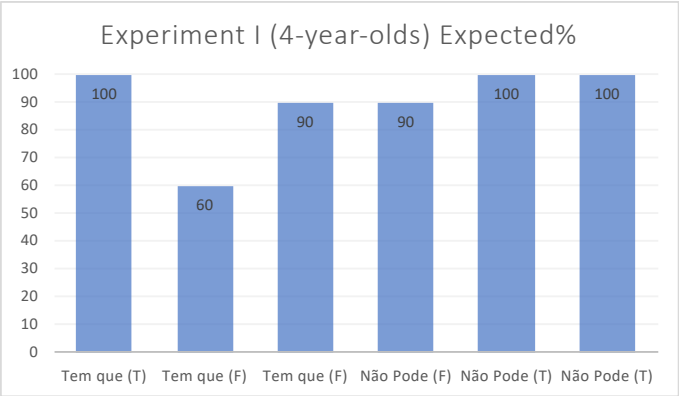
[36] Lunguinho (2004) studied the spontaneous speech of two children between 1;02 and 4;11 years of age. He found that the modal *dever* in the children's grammar only expresses epistemic modality. This is another reason why it would have been problematic to use *dever* in the test sentences as a deontic modal.

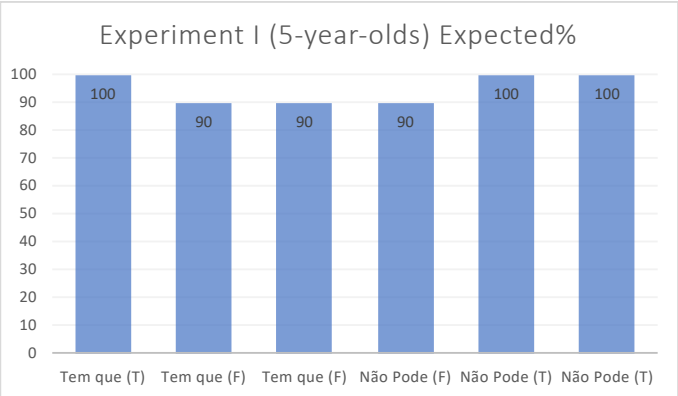
Imaginação" (museum), children's bookstores, activity centers ("Grapali Brinquedoteca", "Terra do Nunca Brinquedoteca" and "Play Space") and also children of parents that I have a friendship with in Brazil. For analysis purposes, children were divided in four groups by age (each group had 10 children): 4;0 to 4;11 years of age, 5;0 to 5;11 years of age, 6;0 to 6;11 years of age and 7;0 to 7;11 years of age. Fifteen adults were tested as well.

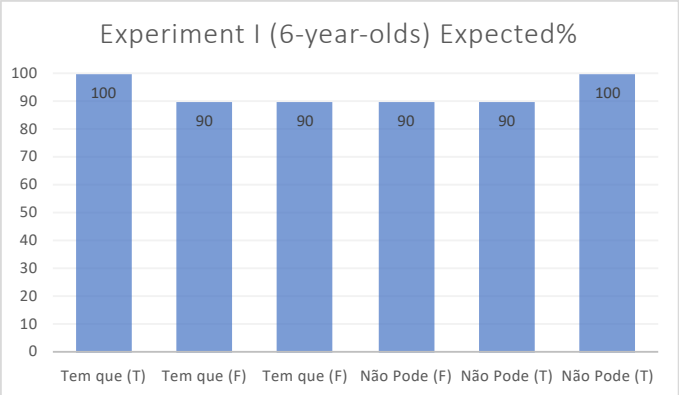
3 Results and Discussion

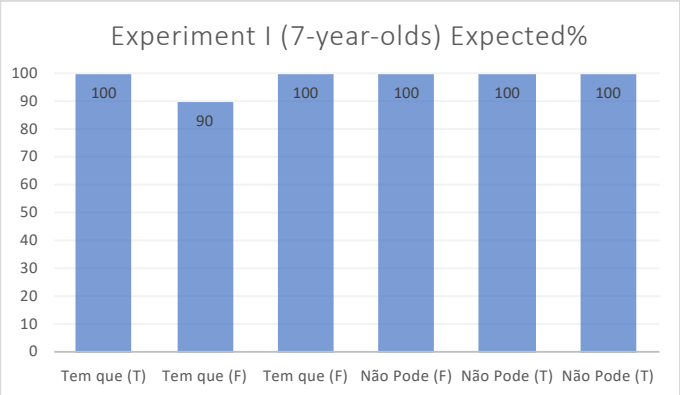
The total number of children interviewed, including the ones that had to be excluded from the analysis, is forty-two. The number of children included in the analysis was forty. One child was excluded because he was too old (8-years-old). Another one (4-years-old) was excluded because he did not want to answer the questions. Children who answered the training items wrongly were corrected until they demonstrated that they understood the experiment. Filler items were answered as expected by the children who were included. The results for each age group are presented in the graphs below. Each graph provides the percentage of responses that matched the expected answers for each test item (in the order they were presented in Table 1, above):

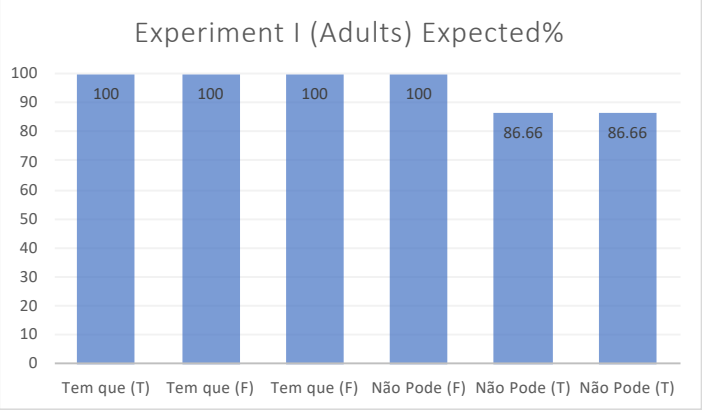
Graph 1: Percentage of expected answers, Experiment I (N = 10, for each child group bar chart, N = 15, for the adult bar chart)







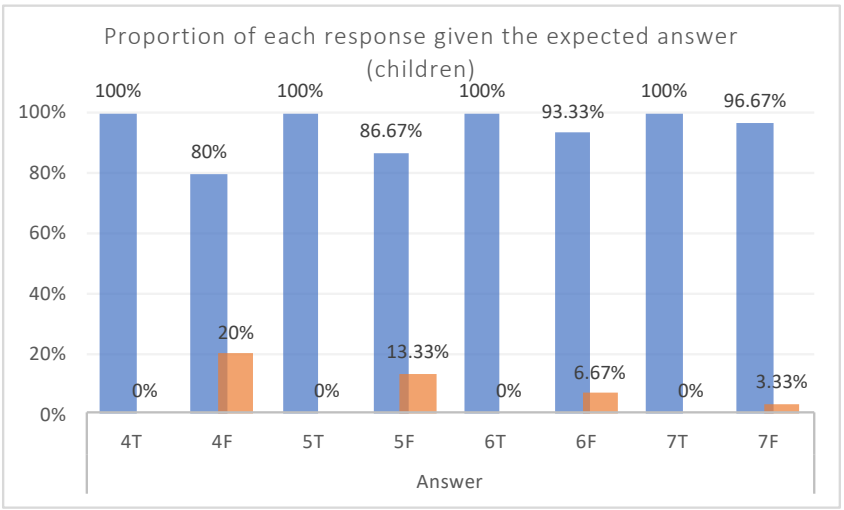




As the graphs suggest, children behaved adult-like (supplying more than 85% of expected answers for each item), with the exception of 4-year-olds on the second item with *tem que*. Notice, though, that this happened for only one item, and that the group performance was adult-like otherwise. It could be argued that the rate of expected answers was only 60% for this

particular test sentence because it was the first 'false' test sentence which was presented to the children. They had to exclude, for the first time in the test, the definite reading of the null subject. For 4-year-olds, false test sentences could be the hardest items in the whole experiment and these sentences could have caused them, in their first judgement, some confusion. In fact, the statistical analysis shows an effect of the answer (whether it was true or false); children were more likely to give the wrong answer when the expected answer was false (Wilcoxon signed-rank test: $Z = 2.78$, $p < 0.01$), as can be visualized in the following bar chart (Graph 2). The y-axis quantifies the proportion of true and false answers that were given by the children. The x-axis contains the ages (4-, 5-, 6- and 7-year-olds) and an indication of the expected answer (F(alse) or T(rue)). As we can see, mistakes were *only* made when the expected answer was F (mistakes are indicated by orange, while corrected responses are indicated by blue) for all age groups.

Graph 2: Bar chart showing an effect of the answer type (children)



The non-parametric Kruskal-Wallis test was conducted to see if there is a difference between children's age groups. Children were not compared to adults, since the level of accuracy

of 7- and 6-years-old was already adult-like, with a percentage of unexpected answer of only 3% for the second group and 1.6% for the first group. The results of the Kruskal-Wallis test show that there is no significant difference between any age group (Chi-square = 0.734, p = 0.8651). In other words, in general, all the groups of children provided expected answers and there is no significant difference between them. No difference between the conditions *tem que* and *não pode* was found, except for 4-year-olds, as the results of the Wilcoxon signed-rank tests show:

Table 2: Difference between conditions, by age group

Age Group	Wilcoxon signed-rank (2-tailed)
7-year-olds	p = 0.317
6-year-olds	p = 1.000
5-year-olds	p = 0.157
4-year-olds	p = 0.046

The difference between conditions for 4-year-olds was caused by 60% of expected answers for the second item of the condition *tem que*.

So far we evaluated differences between groups. Another question is whether children show sensitivity to the adult grammar within groups. In order to answer this question, Wilcoxon signed-rank tests were performed for each age group. If children show sensitivity to the adult grammar they should have said 'yes' only when that was the expected answer. This conclusion can be drawn if the Wilcoxon signed-rank test detects a difference between 'yes' answers to true and false test sentences. The rate of 'yes' answers to true and false sentences is shown in the graph below:

Graph 3: Bar chart showing difference between 'yes' answers to true and false sentences

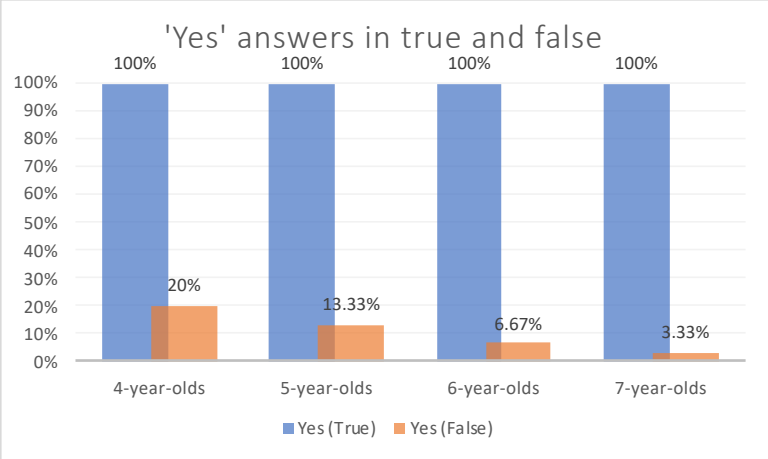


Table 3: Difference between 'yes' answers to true and false sentences by age group

Age Group	Wilcoxon signed-rank (2-tailed)
7-year-olds	p = 0.002
6-year-olds	p = 0.003
5-year-olds	p = 0.004
4-year-olds	p = 0.004

As can be seen in Table 3, a significant difference between 'yes' answers to true and false sentences was detected for all the groups. The results support the hypothesis that all age groups show sensitivity to the adult grammar.

4 Conclusion

By the results presented in the last section, we can conclude that 4-year-olds are sensitive to the adult grammar. That is to say, by this age, children correctly reject the definite reading of the null subject in structures like (3) and correctly assign the generic reading to the null subject.

(3) Nessa escola não *e* pode escovar os dentes depois de comer. *BP*
In.this school not can brush:INF the teeth after of eat:INF
'In this school one cannot brush one's teeth after eating.'

Of course the readers might be concerned that children performed well in the experiment not because they have the knowledge that in BP the null subject of impersonals should be generic, but because the definite subject was not salient enough. In our stories the character "Joaquim" was always the possible definite subject. If it was the case that this character was not salient enough in the stories, children would have no choice but to choose the generic reading: the one that would be always the expected answer, regardless of their grammar. Therefore, a failure in the methodology would explain why children performed well in the experiment.

Nevertheless, it doesn't seem to be the case that "Joaquim", the possible definite subject, was not salient enough in the stories. If we look at the slides from Experiment 1 in the Appendix, we see that Joaquim was the last character to appear and be mentioned before the slide in which Elmo is about to say the test sentence. That was the case even for the fillers and training items. Besides that, Joaquim was salient because he was either the only character that would be the exception to a general rule, or the only character to which a rule would apply. Going back to Table 1, for the impersonal sentence translated into English as, "In this school one has to bring

pets," Joaquim is the only one in the school that doesn't bring a pet. For the sentence, "In this school one has to eat dessert for lunch," Joaquim is the only student that has to eat dessert for lunch, while all the other students eat regular food. In the story for the sentence, "In this school one has to play in the classroom," while all students play outside, Joaquim is required to play in the classroom. For "In this school one cannot study in the morning," all students are required to study in the morning, while Joaquim has to take a nap during the daytime. In the story for the sentence, "In this school one cannot brush one's teeth after eating," the students receive a weird recommendation from a dentist that they shouldn't brush their teeth after eating, but Joaquim, who has big teeth, is the only student that has to brush his teeth after eating[37]. Finally, in "In this school one cannot turn in homework," while all the children are not required to turn in their homework assignments, Joaquim is the only one that is required to do that. In sum, Joaquim is salient enough in the context: he not only appears and is mentioned in all the slides that precede the ones that prepare the child for the test sentence, but he is also the most noticeable character in the stories due to his deviance to the general rules or to the fact that the rules only apply to him, sometimes.

Not being aware of any other confounding factor that could explain why children performed well in this experiment, it can be concluded that children as young as 4-year-olds reject the definite reading of null subjects in impersonal sentences in BP. These results, together with the research showing that Brazilian children acquire morphological agreement by the age of 3 (Magalhães 2006), suggest that Brazilian children already know that they are acquiring a partial null-subject language by the age of 4.

---

[37] The rules presented in the stories from Experiment 1 were explained to the child to be absurd and inappropriate if applied to the real world before I started each test. The effect was that the stories would sound amusing to the children and they didn't seem bored while watching them. Sometimes children would request me to play the stories over again to them.

Chapter 5: Experiment 2

0 Introduction

This chapter discusses the second experiment that was conducted with children acquiring Brazilian Portuguese (BP) as their native language and with adults (as a control group). While the first experiment was designed to test whether Brazilian children understand the null subject in impersonal sentences as having a generic reading, the second experiment was designed to see in which contexts children and adults allow null impersonals in BP.

As discussed in Chapter 3, null impersonals in BP with a generic reading were well-accepted when having a fronted adverb and a deontic modal (ADV+D), in an online survey in which adult BP-speakers participated. On a scale from 1 to 5, 5 being the most acceptable sentences, the average score for structures like (1) was 4.77, which was the highest average score in the whole survey. Particularly, the average score for sentences like (1) was higher than for structures with a generic null subject that only had a fronted adverb with no modal (ADV) (2.18). Generic null impersonals without an adverbial and no modal (NoADV) also received a low average score (2.64). As explained in Chapter 3, a factor that could be responsible for the low average score that sentences in these two conditions received was that there was only one sentence for NoADV and ADV, while all the other conditions had two test sentences each. Therefore, it was important to test these structures again in adults.

(1) Nessa escola *e* tem que PRO trazer animal de estimação. *BP*
In.this school has that bring.INF pets
'In this school one has to bring pets.'

The experiment to be reported was also designed to assess in which structures children allow null impersonals. Recall that in Experiment 1 (Chapter 4), children as young as 4;0 have knowledge that null subjects in impersonal structures have a generic reading. However, Experiment 1 did not test whether children know the restrictions on null impersonals (i.e., the requirements that an overt marker of genericity should be present in the structure).

Our theoretical approach predicts that null impersonals will be acceptable if the structure has a fronted adverbial and a deontic modal (ADV+D condition), but not if it has a fronted adverbial an epistemic modal expressing possibility (ADV+EP)[38]. Recall that as the generic operator is a modal of necessity, deontic modals are associated with genericity, but not modals of epistemic possibility. In the experiment to be reported here, four structures were tested: NoADV (No adverb), ADV (Adverb and no modal), ADV+D (Adverb + deontic modal) and ADV+EP (Adverb + epistemic modal).

1 Methodology and Material

The second experiment consisted of a "*Who said it better?* Task" (i.e., Felicity Judgement Task), particularly a version of the Puppet Competition Task employed by Foppolo et al. (2012). In the experiment, the participants had to watch a sequence of Power Point animations. At the end of each story, they would see a screen with two characters, Elmo and Cookie Monster. In order to see in which contexts null impersonals were accepted, one of the characters would use a

[38] The fronted adverbial was important before testing the hypothesis whether impersonal structures in BP need to have [Spec, TP] filled to check the EPP. As shown in the results of the survey presented in Chapter 3, impersonal structures in BP do not need to have an element in [Spec, TP]. When Experiment 2 was designed, I followed Holmberg (2005, 2010a, b) in assuming that elements such as fronted adverbials are required in impersonal structures with a null generic pronoun in order to satisfy the EPP. However, after the results from the survey, we know that fronted adverbials are not necessary in these structures.

sentence with the impersonal clitic *se* (i.e., SE-pronoun)[39] and the other one produced a comparable sentence with a null subject. The participants had to judge which one said the sentence in "the best way". The participants were told that there were always four choices: Elmo could be right, Cookie Monster could be right, both of them (if they had no preference for the null subject or the SE-pronoun), or none of them (if both structures were unacceptable). Let's see below an example of one of stories (see Appendix C for all the stories that were presented):

Slide 1 A weird rule in this school is that the students cannot eat bananas for lunch.

[39] Although impersonal *se* is pronounced in the same way as the reflexive *se*, I take them to be different lexical elements. As we can see in the sentences below, reflexive *se* can agree with the verb in the third person plural, but not impersonal *se*:

(i) Os homens se barbeiam.
The men REFL shave: 3PL
'The men shave themselves.'

(ii) *Nessas escolas se comem doces.
In.this schools IMP eat:3PL sweets
'In this school they eat sweets.'

(iii) Nessas escolas se come doces.
In.this school SE eat:3SG sweets
'In this school one eats sweets.'

Slide 2 Bruno almost forgot this rule and was about to eat a banana.

Slide 3 But then he remembered it and ate an orange instead.

Slide 4 Now let's see if there is a character that says in the best way something that happened in this story.

Slide 5: **Elmo:** Nessa escola não *e* pode comer banana.

In.this school not can eat.INF banana[40]

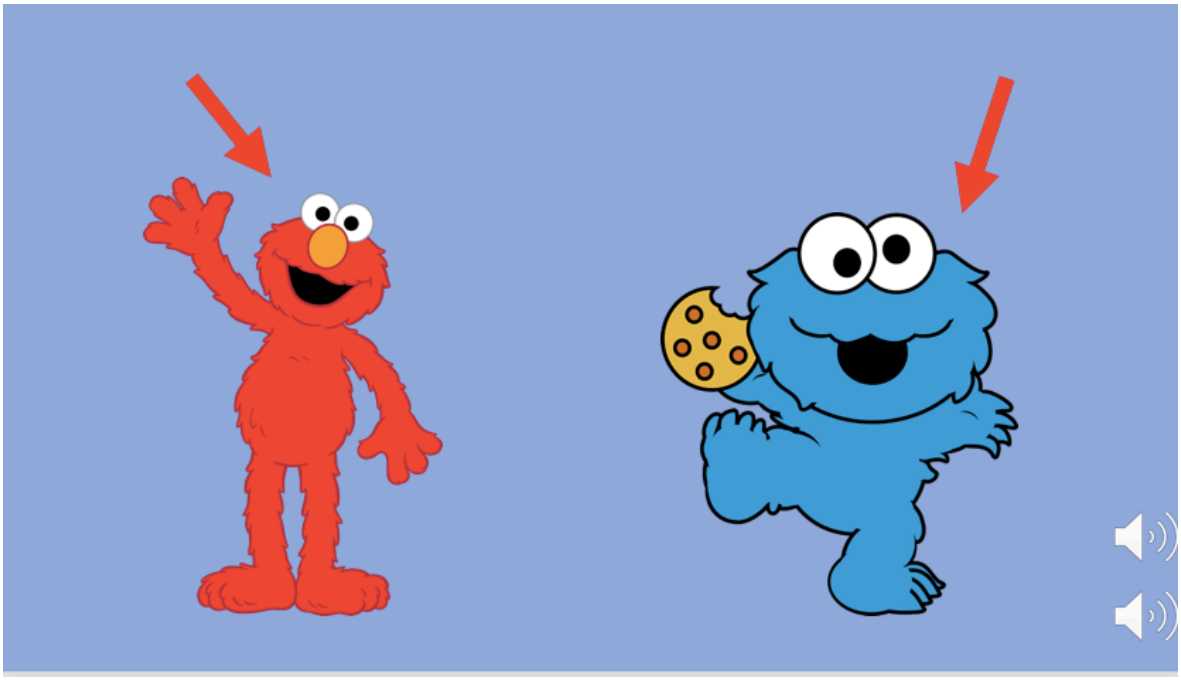
'In this school one cannot eat bananas.'

Cookie Monster: Nessa escola não *se* pode comer banana.

In.this school not SE can eat.INF banana

'In this school one cannot eat bananas.'

[40] Throughout this thesis, the modal *poder* is translated as *can* when it has a deontic interpretation, but it is translated as *may* when it has an epistemic interpretation. This is because *can* without negation disfavors the epistemic reading in English (Hofmann 1976).



In particular, for the example above, as the test sentence had a modal with deontic interpretation (*não pode,* 'cannot'), we expected that the participants would accept both the null pronoun (the sentence uttered by Elmo) and the sentence with a SE-pronoun (uttered by Cookie Monster). Recall, as explained in Chapter 3, that *se* can appear with a deontic modal in null impersonals for BP-speakers. Therefore, the answer "both" was the one that was expected in this case.

The experiment was composed of four pairs of training items, four pairs of test sentences and six pairs of fillers randomly presented between the test sentences. For a list of all the sentences and stories employed, see Appendix C. In this experiment, the sentences always would come in pairs because the participant had to choose between a version uttered by Elmo and/or Cookie Monster or neither of them.

The table below shows the generic impersonal structures that were tested: their conditions with their respective sentences. Each condition had two test sentences that were identical, except for the fact that one of them had a null subject and the other one a SE-pronoun (this optionality is indicated by (se) in the table).

Table 1: Test sentences, Experiment 2

Conditions	**Test sentences**	**Expected Answer**
No fronted adverb (NoADV)	(Se) come almoço. (SE) eat:3SG lunch 'One eats lunch.'	neither
Fronted adverb (ADV)	Nessa escola (se) traz brinquedo. In.this school (SE) bring:3SG toy 'In this school one brings toys.'	'se'
Fronted adverb and deontic modal (ADV+D)	Nessa escola não (se) pode comer banana. In.this school not (SE) can eat:INF banana 'In this school one cannot eat bananas.'	'both'
Fronted adverb and epistemic modal (ADV+EP)	Nessa escola (se) pode ganhar uma medalha. In.this school (SE) may earn:INF a medal 'In this school one may earn a medal.'	'se'

We expected the condition NoADV to be judged as unacceptable. In the version *without se*, there is no modal in the sentence to act as an overt marker of genericity. Regarding the version *with se* of the sentence, as discussed in Chapter 3, *se* in the beginning of a sentence appears after the verb rather than before it in BP.

We expected that participants would prefer the insertion of a SE-pronoun for the structure ADV. This is so because there is no overt expression of genericity in these sentences in the version *without se* (there is no deontic modal).

For the condition ADV+D, we expected that the null impersonal would be accepted, along with the version with a SE-pronoun. In the survey reported in Chapter 3, this structure had an average acceptability of 4.77 in Test 1 (without se).[41] We expected that the participants would accept both SE and a null subject in this condition.

ADV+EP were not tested in the survey reported in Chapter 3. But in a survey that I conducted previously (Bertolino 2017), null impersonals with epistemic modals received an average of 3.3. Therefore, given the empirical facts, we expected that the null subject would be accepted less frequently than the equivalent structure with the clitic *se* in the condition ADV+EP. Given the theoretical framework being adopted, acceptability of null impersonals in the condition ADV+EP with a modal relation indicating possibility is not expected, but if it occurs it might be explained by the fact that participants assigned the deontic necessity reading to the modal.

2 Participants

For the second experiment, I interviewed forty children acquiring BP as their native language[42]. It is important to notice that all these children were different from the ones tested in the first experiment. They ranged in age from 4;1 to 7;8 (mean age = 5;5). Children were recruited in the same places where I recruited children for the first experiment: museums, children's bookstores, activity centers etc. For analysis purposes, children were divided into four

[41] Recall that in Test 2 (with se), *tem que* was used instead of *não pode* for the condition ADV+D. It was suggested that speakers disfavor the co-occurence of *se* with *tem que,* since an average of 3.27 was obtained in Test 2, which was low compared to the average of 4.77 obtained for the same structure without *se*.

[42] No child had to be excluded from this experiment.

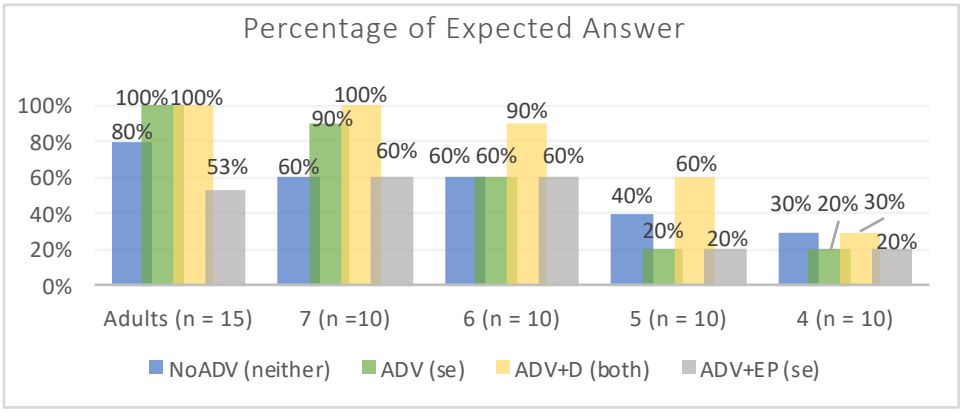
groups according to their age (each group had 10 children): 4;0 to 4;11 years of age, 5;0 to 5;11 of age, 6;0 to 6;11 years of age and 7;0 to 7;11 years of age. Fifteen adults were also tested.

3 Results

Three of the children (4- and 5-year-olds) had initial problems understanding the training items. I went over them with these children until they demonstrated a clear understanding of the task, by being able to judge that sometimes Elmo would say the sentence in the best way, sometimes Cookie Monster would say it in the best way, sometimes both of them and, in some cases, neither of them (each of the four training items covered one of the four possibilities). None of the participants had problems with the filler items.

The results for each each group are shown in the graphs below, with the four conditions presented in Table 1 represented by separate bars. The expected answers are shown in the legend of the bar graph: they were the answers that we were expecting given the results of our previous survey and the theoretical assumptions that we are adopting, as explained above. For ADV+D, we expected that both sentences would be rejected ('neither'), for ADV, *se* was expected ('se'), for ADV+D, we expected the participants to choose both sentences ('both'), for ADV+EP, *se* was the expected answer ('se').

Graph 2: Percentage of expected answers, Experiment 2



The Friedman test was conducted to see if adults performed statistically different in the four conditions. No significant difference was found between conditions (Chi-square = 6.84, p = 0.0772). It is important to keep in mind, though, that with the limitations of the experiment (i.e., each one of the conditions had just one item) and with the fact that only fifteen participants were tested, significance would only be reached with a large difference between responses to each condition.

Let's now analyze the results for children. First of all, Mann-Whitney tests were performed to determine if 7-year-olds behaved significantly different from adults: no difference was found, as seen in Table 2. Then a Generalized Linear Model was conducted comparing the performance of different children's groups, considering an interaction between the independent variables condition and age[43]. No interaction between these two variables was found. Also, no significant difference between conditions was found. But considering the general performance of the groups, 4-year-olds and 5-year-olds exhibited worse performance than 7-year-olds which were taken as the baseline (for both groups: $Z = -2.721$, $p < 0.01$). 6-year-olds were statistically identical to 7-year-olds ($Z = -1.454$, $p = 0.146$).

Table 2: Values of U and p-values for 7-year-olds and adults (per condition)

Conditions	**Values of U**	**P-values (2-tailed)**
NoADV	60	p = 0.424

[43] *R formula: Accuracy ~ Conditions * Age + (1 | Participant).*
R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL
https://www.R-project.org/.

Conditions	Values of U	P-values (2-tailed)
ADV	67.5	p = 0.697
ADV+D	39	p = 0.976
ADV+EP	39	p = 0.802

In order to test whether children have an adult grammar, answers to the conditions ADV vs. ADV+D were compared for each age group. The conditions ADV and ADV+D were chosen because adults gave the expected answer 100% of the time for these conditions, but not for the other ones. I calculated the probability of children answering 'both' for the conditions ADV and ADV+D. The rationale is that if children performed adult-like in the conditions ADV+D and ADV, they should have said 'both' significantly more often for the condition ADV+D than for the condition ADV (i.e., likewise, the probability of choosing 'se' for both conditions could also be calculated rendering the same results). Wilcoxon signed-rank tests were performed for the comparisons. As we can see in Table 3, 4- and 5-year-olds did not differentiated the conditions ADV+D and ADV from each other, but the other age groups did.

Table 3: Probability of saying 'both' for the conditions ADV and ADV+D

Age Groups	Percentage of 'both'	P-values (2-tailed)
7-year-olds	ADV (10%), ADV+D (100%)	p = 0.003
6-year-olds	ADV (10%), ADV+D (90%)	p = 0.005
5-year-olds	ADV (30%), ADV+D (60%)	p = 0.188
4-year-olds	ADV (20%), ADV+D (30%)	p = 0.317

It can be concluded from these results that 7- and 6-year-olds behaved adult-like, but the same cannot be concluded regarding 5- and 4-year-olds.

4 Discussion

First I will discuss the results for adults. The structure NoADV (neither) received unexpected answers 20% of the time. Out of fifteen participants, two of them said that both characters described the story in the "right" way and one of them accepted the structure with *se* ("Se come almoço"). The most complex aspect of the test might be considered to be the item whose expected response is 'neither' (NoADV), if we think that most of the time participants are making comparisons between sentences and expecting to select the utterance (or utterances) that best suits the story. Therefore, the option 'neither' might not be mentally considered most of the time.

It could be plausibly argued, though, that the alternative 'neither' is not out of consideration by the participants, since they had no problem choosing this alternative in the training item that had this expected answer. However, two aspects need to be considered here: first of all, when participants started the training section, they were just told that "sometimes both of the characters say something wrong or something that sounds weird", in such a way that this information (that 'neither' is possible) would be still "fresh" in their minds, but possibly it was not anymore when they were giving answers to the actual test. The second aspect to be considered is that the task of judging the pair of sentences presented as a training item and the one presented as a test sentence (NoADV) was of a different nature, even though the expected answer was 'neither' for both of them. The pair of sentences for the training item were both clearly factually wrong: the story was about Mariana and Joaquim at the school playground. Mariana considers playing in the sandbox, on the slide or on the swing. *She ends up playing on the swing*. Elmo and Cookie Monster then say:

(2) **Elmo:** A Mariana brincou na areia.

The Mariana played in.the sand

'Mariana played in the sandbox.'

Cookie Monster: A Mariana foi no escorregador.

The Mariana went on.the slide

'Mariana played on the slide.'

Both sentences from this training item are obviously false to any participant that paid attention to the story. They had to judge the *truth value* of the sentences, given a certain context. On the other hand, the pair of sentences for the condition NoADV required the participants to judge the *grammaticality* of the sentences, a task qualitatively different from the training item. A sentence is either *true* or *false,* but the grammaticality of some sentences might fall in a doubtful area, while other sentences are categorically grammatical or ungrammatical.

We should keep in mind, though, that three out of fifteen non expected answers is a small number to make any sort of speculation about; due to the complexity of the task, in particular the need to choose between four possible alternatives, it seems reasonable to adopt the adult response level of 80% expected responses as a comparison point for children's responses on this item.

For the conditions ADV (se) and ADV+D (both), adults provided 100% expected answers. As mentioned above, the expected answer was 'both' for the condition ADV+D because

it was the only condition in which we were expecting a null subject to be accepted, although the sentence with a clitic *se* is still acceptable.

For the condition ADV+EP ('se'), eight out of fifteen participants chose the expected answer with the clitic *se* (53.33%). Five adults chose the option with the null subject (33.33%) and the other two adults chose 'both' (13.33%). Given the theoretical framework adopted, this is a puzzle. Our approach would only predict null impersonals if the participants interpreted the modal as deontic in the test-sentence that was used, making it equivalent to the condition ADV+D (in which the participants chose the option 'both' 100% of the time).

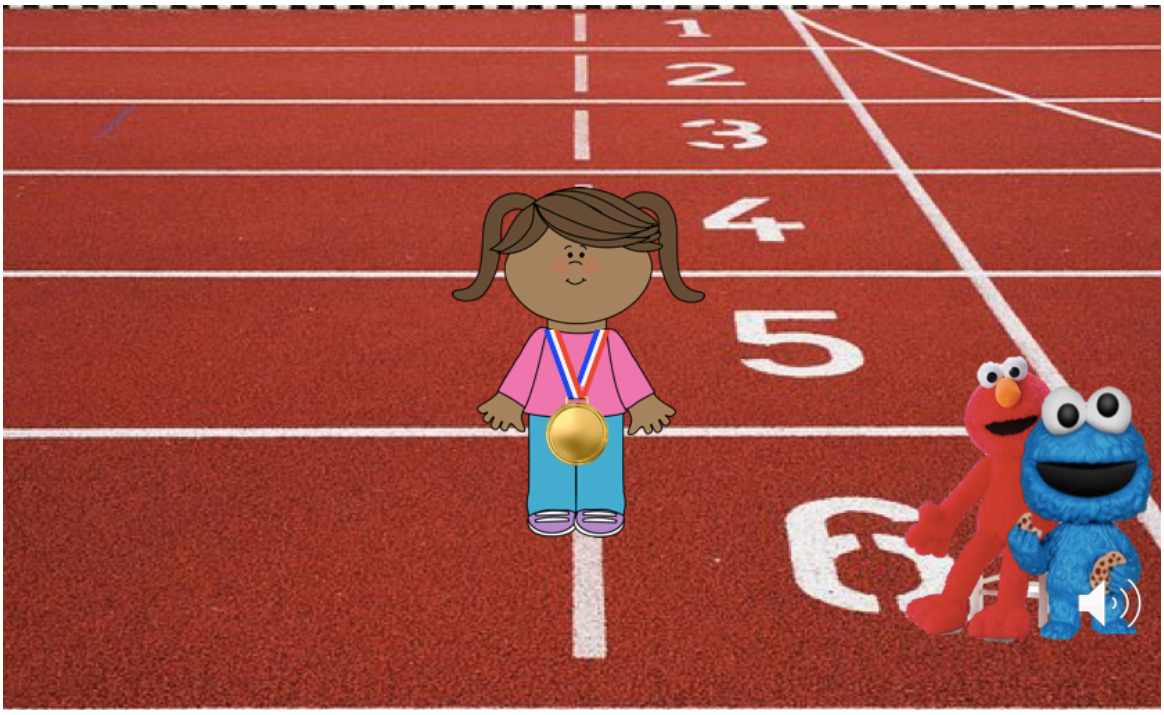
(3) Nessa escola (se) pode ganhar uma medalha. *ADV+EP*
In.this school (SE) may earn.INF a medal
'In this school one may earn a medal.'

The intended interpretation for sentence (3) was the one in which the modal had the epistemic reading: in the school, there is a possibility to earn a medal. Below is the narrative for the sentence:

Slide 1 One day, the school was having a competition to see who could run faster. Joaquim almost wins the competition…

Slide 2 But, Mariana ended up winning it.

Slide 3 Mariana earned a medal.



Slide 4 Now let's see if there is a character that says in the best way something that happened in this story.

Slide 5: **Elmo:** Nessa escola *e* pode ganhar uma medalha.

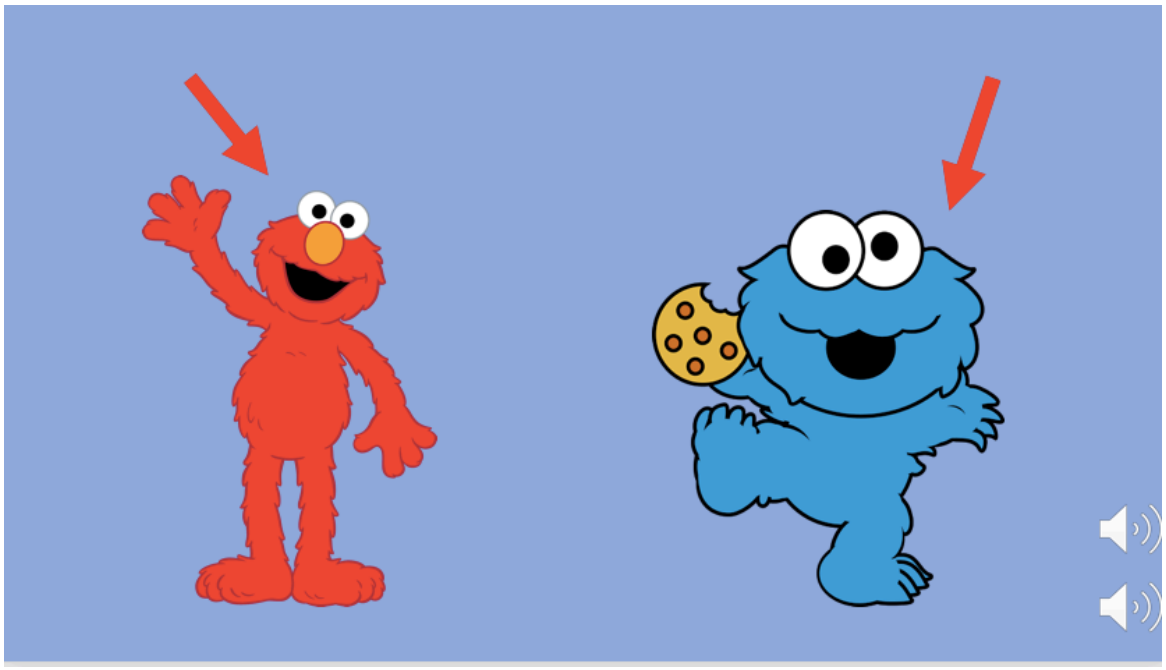
In.this school may earn.INF a medal

'In this school one may earn a medal.'

Cookie Monster: Nessa escola *se* pode ganhar uma medalha.

In.this school SE may earn.INF a medal

'In this school one may earn a medal.'



It could be argued that some participants had access to the deontic reading in the sentence for the condition ADV+EP: in the school one is *allowed* to earn a medal. However, although the modal *poder* is potentially ambiguous, this is not likely because, as explained in the

Chapter 4, *poder* is very often associated with the epistemic reading when it is not preceded by the negation.

Also, if some participants had interpreted the modal as deontic, we would expect that they would always provide the answer 'both' as they did for the condition that undoubtedly had a deontic modal (ADV+D). But only two of the participants who did not provide the expected answer to the condition ADV+EP said 'both': the majority in this group of participants preferred only the sentence with the null subject.

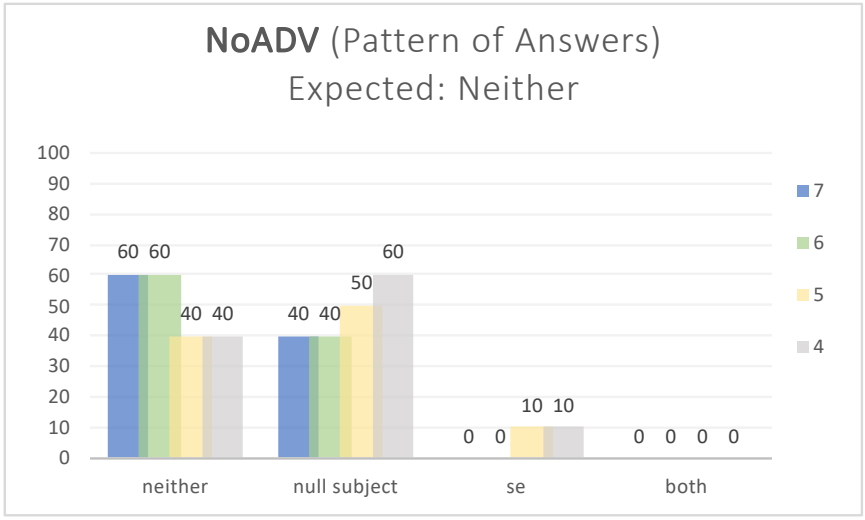
It is also possible in principle that participants did not provide the expected answer for ADV+EP because *se* cannot combine with the modal *poder*. In Chapter 3, it was discussed that a similar restriction arises for *tem que*. Nevertheless, in my own variety of BP, *se* can be combined with *poder*.

In sum, *it is certain that adults differentiate the condition ADV+D from ADV+EP, given the distinct nature of the response pattern provided for each condition*. It is not easy to answer, though, what caused the mixed responses for the condition ADV+EP, as more than one interpretation is possible. We could argue that maybe speakers have different judgments related to the condition ADV+EP: some of them accept null impersonals in this conditions, while others do not. There is another possible interpretation: as the difference between the conditions ADV+D and ADV+EP is quite subtle, the participants seemed to have recognized some difference

between these conditions, but were unsure about which judgement they should provide for ADV+EP[44].

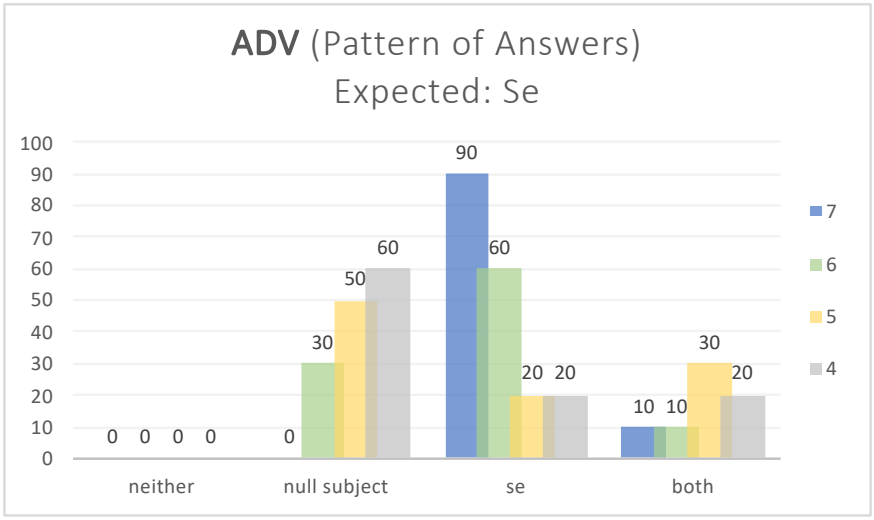
Now I turn to the discussion about children's results. Children's erroneous choices were analyzed. The graphs below show that young children (mainly 4-year-olds) tend to choose the sentence with a null subject when giving the wrong answer. They rarely chose wrong responses with *se*: 4-year-olds and 5-year-olds together only chose *se* three times as a wrong answer, during the entire test. In fact, as we saw in Graphs 4 and 6, 5- and 4-year-olds rarely choose *se* even as a correct answer. It seems that the acquisition of impersonal *se* is somehow delayed. One child that I tested (5-years-old) explicitly asked me what *se* was, demonstrating that he seemed to lack knowledge of the form.

Graph 3: NoADV - pattern of answers by age group (N = 10, for each child group bar chart)

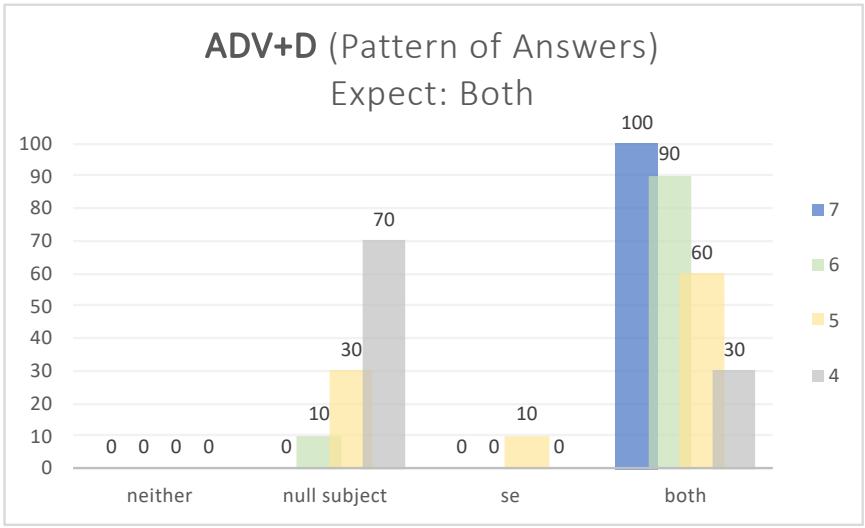


[44] I am not suggesting that the responses that the participants provided for the condition ADV+EP were influenced by their acceptability of the null subject in the condition ADV+D *in the actual test*, because the condition ADV+EP preceded the condition ADV+D in the experiment (see Appendix C for the order in which all the sentences were presented). I am suggesting, rather, that their subconscious knowledge that null impersonals are allowed when we have a fronted adverb and a deontic modal in BP might have caused them some confusion to recognize that the null subject is not possible when there is an epistemic modal, since the small difference between the two conditions might be too complex to be promptly distinguished in an experimental setting.

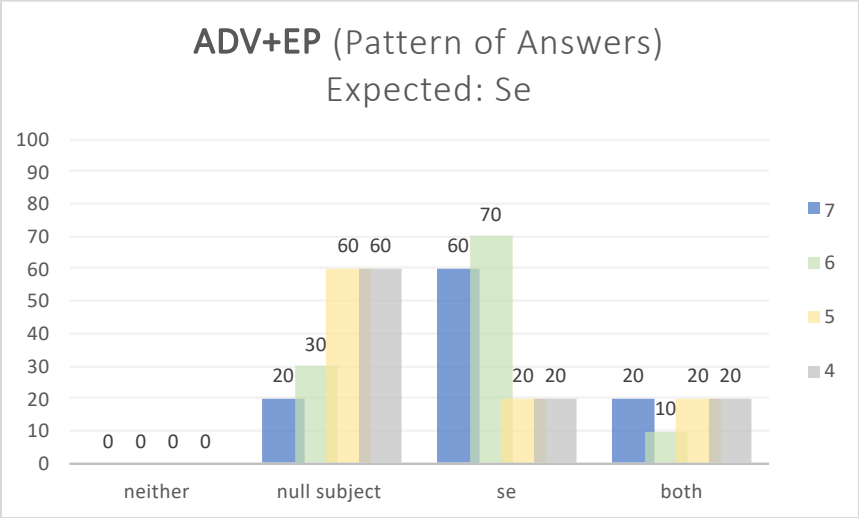
Graph 4: ADV - pattern of answers by age group (N = 10, for each child group bar chart)



Graph 5: ADV+D - pattern of answers by age group (N = 10, for each child group bar chart)



Graph 6: ADV+EP - pattern of answers by age group (N = 10, for each child group bar chart)



According to Assis (2017), *se* is being less and less used as a strategy to form impersonal sentences. In 1970, there was a balance between the percentage of the options used in spoken BP to form impersonals with a generic interpretation (i.e., null impersonal, impersonals with *você* and *se*). In the presence decade, the use of *se* has declined from 16% to 3%. In contrast, the use of null impersonals has declined from 23% to 12%.[45] It is likely, therefore, that *se* is not largely present in the input BP-children are receiving.

Silva (2017) hypothesizes that impersonal *se* is learned at school and would not be part of the initial grammar of a child acquiring BP. She shows that the form appears first in written material instead of spoken language. It was almost absent (0.23%) in a corpus of the speech of two children between the ages of 2 and 5 (200 transcriptions were analyzed). In a corpus of the writing of 7- to 10-year-olds (approximately 100 writing samples), the form was also rare (2%).

[45] The pronoun *você* appears as the preferred strategy to form impersonals with generic interpretation in the data of the author. For details, see Assis (2017: 116).

When Silva analyzed 120 textual productions of teenagers in a college entrance exam whose grades were above average, there was 13% use of impersonal *se*. In the speech of adult native speakers of BP who were considered to use the so called "Norma Urbana Culta" in São Paulo (NURC/SP) (the linguistic norm used by educated and cultured individuals from the São Paulo urban region), the impersonal *se* was used 61% of the time (the author analyzed 40 conversations between two participants and 20 elocutions in a more monitored speech style between 1970 and 1980)[46]. In 480 political news articles collected from a newspaper from São Paulo, Folha de S. Paulo, from 1980 to 1990 and 2000 to 2010, the occurrence of impersonal *se* was 24%. The author concluded that the occurrence of impersonal *se* depends on schooling, but it was not necessarily limited to being associated with the written register (given the high percentage of use of the form by NURC speakers in the oral register).

Silva's study shows interesting data that seem, to some extent, to be in agreement with our findings. According to MEC (Ministry of Education in Brazil)[47], 6-year-olds have their first contact with education in its institutionalized sense. At this age their literacy development officially starts in their everyday life with the introduction of a new culture: the school[48]. In our experiment, it was not until children reached the age of 6 that they seemed to perform adult-like and exhibited less preference for the null subject when it was not chosen by adults.

---

[46] The high percentage of impersonal *se* in the data analyzed by Silva in contrast with the data analyzed by Assis (2017), could be explained by dialectal differences. Assis analyzed data from São Paulo, Rio de Janeiro and Minas Gerais. Silva analyzed data only from São Paulo.

[47] http://portal.mec.gov.br/index.php?option=com_docman&view=download&alias=4845-a-crianca-6anos-a-linguagem&Itemid=30192, accessed on January 13, 2019.

[48] Schooling and the formal contact with writing starts when children reach the age of 6 in Brazil. But it could be said that children usually have some earlier contact with written language, through the reading of children's books by parents and teachers at daycare centers. At this point already, impersonal *se* might be more present in the input, although institutionalized education might make children aware of the existence of the form.

5 Conclusion

In conclusion, after running the statistical analysis, we saw that adults performed as expected for the condition (NoADV), in which we expected the answer *none*, for the condition (ADV), in which we expected the answer *se*, and for (ADV+D), in which the expected answer was *both*. However, the same did not happen for the condition ADV+EP: we expected that adults would choose the sentence with *se*, but many adults preferred the sentence with a null impersonal. There are two possible interpretations for this result: (i) some adults, in fact, accept null impersonals when the modal is epistemic; (ii) given their subconscious knowledge that null impersonals are allowed when the modal is deontic and the subtle difference between a sentence with a deontic modal and with an epistemic modal, participants were unsure which answer they should provide for the condition ADV+EP in the experiment. We discussed why the results are not coherent with the idea that the unexpected answers for the condition ADV+EP are explained by assuming that participants actually interpreted the modal as deontic, because the conditions ADV+D and ADV+EP triggered a pattern of responses remarkably different from each other.

Regarding the results for children, in general, 7-year-olds and 6-year-olds behaved adult-like when considering the statistical results. The statistical analysis didn't detect a difference in behavior across conditions for children. 5-year-olds and 4-year-olds, in general, exhibited worse performance than the other groups. Graph 3 to 6 seem to suggest that there is a tendency, among these children, to choose the alternative in which there is a null subject (that is, exclusively the sentence with the null subject, not incorrectly selecting *both*). Taking into account the results presented by Silva (2017), we speculated that impersonal *se* develops with schooling: that would

be why only children who have started their literacy development would choose sentences with *se*.

In contrast to Silva (2017), though, I do not assume that impersonal *se* is not part of the initial grammar. The author proposes that there is an internal grammar that is acquired by natural means, but also another one that would be acquired at school. This last grammar is conservative and based on prescriptive grammar rules. According to Silva (2017), impersonal *se* is only part of this grammar that is not acquired naturally. However, our data provide counterevidence to her conclusion: it is unlikely that it is taught at school that *se* is required when there is no modal in an impersonal sentence, something that adults in our experiment demonstrated to clearly know, as also did 7- and 6-year-old children.[49] When there is a deontic modal, the native speaker of BP knows that a null element is possible in impersonal sentences. This knowledge is not based on prescriptive rules.

I propose, instead, following Costa et al. (2017), that although the school can promote learning of competences that are sometimes not present in the natural process of acquisition (in other words, in the internal grammar), *it can also contribute to the development of grammatical aspects that are already part of the internal grammar of a child*. There are grammatical structures that are acquired early by the child, while others are acquired late (for example, passives (Sim-Sim 1998) and object relatives (Costa et al. 2011)), since they are less frequent in the input and/or more complex. The child at school has contact with a linguistic environment where she is stimulated to use a variety of structures that are part of her language, in a way to develop her argumentative capacity and show her way of thinking in an enriched and well-arranged way.

---

[49] Alternatively to *se*, the pronoun *você* can be used when there is no modal. Experiment 2, though, only contrasted the variants *se* and the null subject.

Costa et al. (2008) show that conjunctions such as *mas* (but), *porque* (because) and *se* (if) are part of the implicit linguistic knowledge of 2-year-old children acquiring European Portuguese. Nevertheless, these conjunctions are only productively used when children reach the 4th year of basic schooling (at age 9).

Experiment 2 was intended to capture the implicit linguistic knowledge of children about structures that allow null impersonals and or that require the use of an overt generic form (in the case, *se*). But the task required children to pay attention to different linguistic forms and judge them. Once children are at school, they move from a mere intuitive linguistic knowledge in which they are unaware of the properties of their language to an explicit knowledge in which they pay more attention to linguistic forms and, therefore, become able to judge the form of sentences (Costa et al. 2017).

In sum, my proposal, more convergent with the empirical results, is that school might be playing two roles: (i) it is promoting the use of a structure that is not widely present in the input, but part of the internal grammar of the speakers - given the subtle judgments that adult speakers are able to provide which are not taught at school; (ii) it is facilitating the process of judging the form of different sentences, since once children are at school they are able to pay more attention to the structure of a sentence and not only to the information that it conveys.

Experiment 1, discussed in Chapter 4, showed that children acquiring BP as young as 4 years of age know that the null subject in impersonal constructions has a generic reading instead of a definite one. Experiment 2, discussed in this chapter, was designed to see whether children know in which structure the null generic pronoun is allowed. While 7- and 6-year-old children did not behave statistically different than adults, 5- and 4-year-olds did. I suggested that these

results emerged because 4- and 5-year-olds have not acquired the clitic *se* leading to an over acceptance of sentences with the null subject.

Chapter 6: Null Subject Parameters and Language Acquisition

0 Introduction

The results of the experiment presented in Chapter 4 have shown that children acquiring BP know that null subjects in impersonal structures have a generic reading. They correctly reject the definite reading of the null subject in this context. Now I turn to the question of how they came to have this knowledge.

In this chapter, I also discuss the acquisition of null subjects across different languages, proposing how children set the null subject parameters. The proposal I present builds upon Holmberg's (2010a) D in T, P in T and ϕ-dependent parameters, which account for a wider range of languages than the initial formulation of the Null Subject Parameter (Rizzi 1982).

The chapter is structured as follows: in section 1, I review previous studies on the acquisition of definite subjects, focusing mainly on English (a non-null-subject language), European Portuguese (EP) (a consistent null-subject language) and Brazilian Portuguese (BP) (a partial null-subject language). I show that although the occurrence of missing subjects in English might suggest that children misset the null subject parameter(s), early parametric analyses to the phenomena have proved problematic. In contrast to English, the data from BP and EP on definite null subjects suggest that children acquiring these languages have knowledge of the distribution of null subjects in their language from very early on.

In section 2, I focus on null subjects in impersonal constructions with generic reading, showing that children acquiring BP and Estonian (both partial null-subject languages) produce such null subjects early in their development: before the age of 3;0. I also discuss the acquisition of impersonal *se* in consistent null-subject languages (Spanish and EP). Children acquiring

Spanish and EP exhibit early use of the impersonal clitic *se,* in contrast to children acquiring BP, whose acquisition of *se* is delayed, as shown in Experiment 2 (Chapter 5). As impersonal *se* is not frequent in the input received by Brazilian children (see Chapter 5), it is not surprising that its acquisition will be delayed compared to children acquiring EP and Spanish, since in these languages *se* is a common strategy to form impersonal constructions.

In section 3, I propose an analysis of how children set the null subject parameters in their language. This analysis is based on the hierarchical organization proposed by Holmberg (2010a). The proposal aims to explain the acquisition of a complex null-subject typology: radical pro-drop languages, non-null-subject languages, consistent null-subject languages, partial null-subject languages and semi null-subject languages. I deduce the initial setting of these three parameters from the Subset Principle. Next, I discuss which evidence in the input could be used by the child to set these parameters across different language types. In section 3.8, I return to the issue of missing subjects in English which was first raised in section 1. I show that missing subjects in English cannot be the result of a misseting of any of the parameters discussed in this chapter (i.e., D in T, P in T and ϕ-dependent parameters). In fact, the initial setting of these parameters corresponds to a non-null-subject grammar, as will be shown in section 3.7. As missing subjects occur in the initial position of the sentences, I propose that they result from subject ellipsis, an option available in the adult grammar of English.

1 Previous studies: Definite Null Subjects

In this section, I shall discuss children's production of definite null subjects cross-linguistically. That is to say, this section does not focus on null subjects in impersonal structures (which I turn to in section 2), but on null subjects with definite reading. This discussion is

important for this thesis because it will provide us with data for our subsequent discussion regarding whether there is evidence of parameter missetting in the child's grammar. In the literature, the phenomenon of missing subjects in non-null-subject languages has been the main instigator for proposals arguing that the child's initial grammar is different from the adult's grammar in the domain of null subjects. Therefore, I start this section by discussing the production of null subjects by children acquiring non-null-subject languages such as English and some of the proposals that try to explain this phenomenon.

1.1 English

English, French and Danish are examples of languages in which the subject cannot be null in the adult language. However, children acquiring these languages are reported to drop subjects during a certain stage in their development, which occurs roughly from 20 to 25 months (Hyams 1986: 65), although the age in which this stage lasts varies from child to child, with some children producing null subjects for much longer. During this stage, non-imperative subjectless sentences alternate with sentences having overt subjects. The following examples from English are taken from Hyams (1986: 65-66) quoting L. Bloom's (1970), L. Bloom et al.'s (1975) and Braine's (1963) corpora (for French, see Rasetti (2000), for Danish, see Hamann and Plunkett (1998)). There is even the presence of minimal pairs such as (1c) & (2c), (1d) & (2d) (produced by Andrew):

(1)a. Throw away.

b. Want more apples.

c. Put that on.

d. Take a nap.

(2)a. Mommy throw it away.

b. I want doggie.

c. Andrew put that on.

d. Mama take a nap.

According to Hyams (1986), children initially set the Null Subject Parameter to 'yes' (null subjects are allowed), and this explains the occurrence of null subjects in non-null-subject languages. In other words, all children start out speaking a language like Italian or Spanish (consistent null-subject language) and later they reset the parameter to 'no' (null subjects are not allowed), if the positive evidence in the language forces the child to do so (e.g., the presence of expletives can be one of the triggers to induce the resetting of the parameter). Hyams (1986) adopts a formulation of the Null Subject Parameter inspired by Rizzi (1982), but with some modifications. For Rizzi (1982), INFL in consistent null-subject languages is [+pronominal], whereas for Hyams (1986), it is AGR which is [+ pronominal]. She argues that AGR is identical to an empty category, since both have intrinsic ϕ-features (person, number and sometimes gender). The empty category to which AGR is associated, according to Hyams (1986), is PRO and grammars vary in whether AGR is or is not PRO. AGR is equal to PRO (AGR/PRO) in pro-drop languages (consistent null-subject languages), but not in non-pro-drop languages (non-null-subject languages). In pro-drop languages, when AGR = PRO, null pronouns are licensed in tensed clauses with definite reading (3a). This null pronoun with definite reading is not licensed in languages like English, since AGR is not equal to PRO (3b).

(3)a. *e* (AGR/PRO) VP *pro-drop*

b. **e* (AGR) VP *non-pro-drop*

Hyams (1986) claims that AGR has the same properties as PRO in pro-drop languages:

(4)i. It may be controlled;

ii. It may be arbitrary in reference and

iii. It is ungoverned.

By assuming that AGR is equal to PRO in pro-drop languages and that children start out with a pro-drop language, Hyams predicts the following phenomena during the period in which the child acquiring English has a pro-drop grammar:

(5)i. Lexical pronominal subjects will be optional.

ii. Null subjects will have a definite reference.

iii. Expletives will be absent (*it* and *there*).

iv. Modals and auxiliaries will be absent.

Lexical pronominal subjects are predicted to be optional (5i) in this account because null subjects are optional in languages like Italian, although, in general, overt pronouns are used in pro-drop languages only when it is desirable to signal emphasis or contrast. Similarly, null subjects are predicted to have a definite reference (5ii) because this is what is found in pro-drop

languages: as already discussed in this thesis (see Chapter 1), the generic reading in pro-drop languages (i.e., consistent null subject languages) cannot be expressed only by a null subject; other strategies are required to express this reading, such as the use of the passive voice, some dedicated impersonal voice or a *si*/*se* pronoun. The reason why Hyams predicts expletives to be absent (5iii) is also clear: pro-drop languages lack expletives and if the child is aware of this fact, it is predicted that her speech will lack these elements.

Why modals and auxiliaries are predicted to be absent from the child's speech (5iv) is a more complicated matter. According to Hyams (1986), in Italian, modals and auxiliaries are in general base generated in VP. If the child has an "Italian grammar" that means that she should analyze modals and auxiliaries with the status of main verbs. According to Hyams, if modals are supposed to be analyzed as main verbs by the child, the child is expected to inflect modals (as if they were verbs):

(10)a. *John is musting go.

b. *John musts go.

c. *John musted go.

(Hyams 1986: 77)

Citing Maratsos (1982: 252), Hyams (1986) points out that children in general do not attach verbal inflections to categories other than verbs. This strongly suggests that they have knowledge of grammatical classes during early stages. Therefore, Hyams concludes, modals (and *be*) are neither analyzed as appearing under AUX (as in non-pro-drop grammars) nor as main verbs (as in pro-drop grammars). Modals and auxiliaries are instead filtered out by the grammar

of the child, which provides no structural description to these elements.[50] They emerge only when the grammar is restructured (when the parameter is reset).

When the grammar is restructured, regular and productive use of overt subjects is followed by the emergence of modals, the progressive and copular 'be'. There could be two triggers to reset the parameter, according to Hyams: (i) expletives, which occur only in non-pro-drop languages and (ii) sentences in which a definite subject pronoun appears in infelicitous circumstances. The hypothesis in (ii) needs clarification: according to Hyams (1986), the Avoid Pronoun Principle dictates that subject pronouns will be avoided except when they are required for contrast and emphasis; assuming that at this point the child has a pro-drop grammar they will also expect null pronouns except for contrast. Once the child learns that in English contrastive and emphatic pronouns are stressed, she will see any subject pronoun which is unstressed as infelicitous, according to the Avoid Pronoun Principle. The child then deduces: if the pronoun is not being used for pragmatic reasons, then it should be necessary for grammatical reasons. At this point the child will conclude that the grammar of English is not an Italian-type grammar.

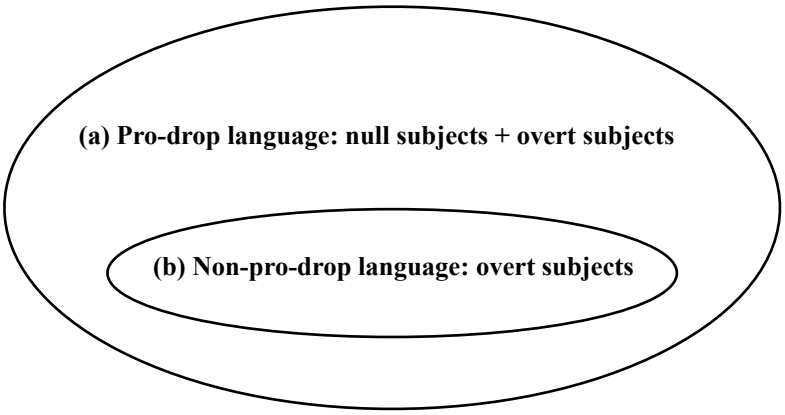
An important contribution from Hyams' (1986) account is that it is consistent with the Subset Principle (Berwick 1982; Wexler and Manzini 1987). According to this principle, initially children should choose the most restrictive value for a parameter (the subset). If the child's initial choice is insufficient to account for the data, children should switch to the least unrestrictive value for the parameter (the superset). Consider the Null Subject parameter. If this parameter only governs the appearance of definite pronouns, English-type languages form the subset or the most restricted value for the parameter, since they only allow overt subjects. Italian, on the other

---

[50] The missing 'be' phenomenon is not as general as the modal phenomenon, as Hyams (1986) herself discusses. Children consistently produce *be* when it cannot be contracted in the adult language (e.g., "here it is"). This is a function of the fact that the uncontractable copula can be analyzed as a main verb, given its inflectional properties.

hand has the least restrictive value for the parameter (the superset), since it allows both null and overt pronouns (although in restricted circumstances, as explained above). This is represented in the figure below:

Figure 1: Null subject parameter and the Subset Principle



However, Hyams'(1986) proposal avoids problems with the Subset Principle by clustering together properties which are not in a subset-superset relation (Hyams 2011). For example, English and Italian are not in a subset-superset relationship with expletives: while English has expletives, Italian does not have them. They are also not in a subset-superset relation regarding the use of modals: English modals constitute a different category than Italian modals. Then, Figure 1 does not represent the full extent of the null subject parameter, and in this way, Hyams' (1986) proposal does not raise a problem for the Subset Principle.

However, Hyams' (1986) approach faces empirical problems. Recall that her approach predicts that expletives will not be produced by English-speaking children while they are dropping subjects. This prediction is inconsistent with Valian's (1991) findings which show that English-speaking children produce expletives while still dropping definite subjects. By

comparing children acquiring English and Italian, Valian (1991) also found that children acquiring English produce far fewer null subjects (30%) than Italian children (70%). This finding is unexpected if young English-speaking children are in a stage at which they have a pro-drop language like Italian. The environments in which null subjects are produced by English-speaking children is also different from the environments where null subjects are produced by Italian children: children acquiring English do not produce null subjects in subordinate clauses and after wh-elements (Valian 1991; Guasti 2002)[51], which is not true for children acquiring Italian (Guasti 1996). That is, English-speaking children only drop subjects when they are sentence initial.

Hyams (1991) provides another analysis of the null subject phenomenon. She proposes that children acquiring languages with a poor inflectional system start out with a Chinese-like grammar (i.e., radical pro-drop language), while children acquiring languages with a rich inflectional system have an Italian-like grammar (i.e., consistent null-subject language)[52]. Hyams develops her proposal based on Jaeggli and Safir's (1989) approach to the null subject phenomenon. According to their proposal, "morphological uniformity" accounts for the possibility of null subjects:

(11) Null subjects are permitted in all and only languages with morphologically uniform inflectional paradigms.

---

[51] See, though, Roeper and Rochrbacher (2000) and Bromberg and Wexler (1995) who found that post-wh null subjects occur with bare verbs in child English in *non-finite* clauses.

[52] As in consistent null-subject languages (e.g., Italian), radical pro-drop languages are also subject to a version of the "Avoid Pronoun Principle" (Kang 2014) . However, there are cases in which overt arguments are preferred in radical pro-drop languages. Differently than consistent null-subject languages, radical pro-drop languages do not use overt pronouns for emphasis, but for other purposes. For example, in Chinese, an overt object is used when it is bound by the matrix subject. See Huang (1984) for discussion.

(Jaeggli and Safir 1989: 29)

In order to have a uniform paradigm, languages should either have an inflectional paradigm which consists entirely of complex forms, as the example in Italian below illustrates (12a), or an inflectional paradigm without any complex form, as the example in Chinese (12b) illustrates. As English does not have a uniform paradigm, it is considered a mixed system, as we can see in (12c):

(12)a. 'parlare' (to talk)

1SG	parl**o**	1PL	parl**iamo**	*Italian*
2SG	parl**i**	2PL	parl**ate**	
3SG	parl**a**	3PL	parl**ano**	

b. 'tánlùn' (to talk) *Chinese*

1SG	tánlùn	1PL	tánlùn
2SG	tánlùn	2PL	tánlùn
3SG	tánlùn	3PL	tánlùn

c. 'talk'

1SG	talk	1PL	talk	*English*
2SG	talk	2PL	talk	
3SG	talk**s**	3PL	talk	

Italian and Chinese have morphologically uniform paradigms, but in different ways. While Italian has different forms for all persons, Chinese has the same form for all persons, since it lacks verbal inflection. English, in contrast, has a mixed system: morphologically complex forms (e.g., talk**s**) coexist with simple forms (e.g., talk). As languages like Italian and Chinese have a uniform paradigm, null subjects are licensed in these languages, but not in languages like English, with a mixed paradigm.

For Hyams (1991), in the early English grammar, null subjects are licensed by morphological uniformity. That is to say, at the point children are producing null subjects, they have not realized that English has a nonuniform system. Once the child realizes English has a mixed system, she should stop producing null subjects. Hyams (1991) shows that the production of lexical subject pronouns interacts with the development of inflectional morphology in English: for Adam (from Brown's Harvard study), inflections are used in more than 70% of the time when he is 2;11. At this exact time, the production of overt subject pronouns increases from 60% to almost 80% and it continues increasing until it reaches adult levels, suggesting that there is a correlation between the development of the inflectional paradigm and the end of the null subject phase.

Hyams (1991) also observes that for children acquiring languages with rich and uniform paradigms, the acquisition of inflections proceeds more quickly than for children acquiring English (Hyams 1983; Weist and Witkowska-Stadnik 1985). This fact is explained by the hypothesis that children tend to make paradigms uniform (Slobin 1973). In other words, the child's initial grammar is uniform. According to this view, if a child is acquiring a paradigm that is already uniform (like Italian), the acquisition of verbal inflections will be less erroneous, since her initial grammar is compatible with the grammar of the language she is acquiring. However, if

the child is acquiring a mixed system, the acquisition of the verbal paradigm will proceed slower and it will be more erroneous, since her initial grammar is incompatible with a mixed system.

From the viewpoint of learnability, assuming that the child's initial grammar is uniform has its advantages: if the child initially assumes that no verbal forms are inflected or that all verbal forms are, positive evidence will tell the child otherwise. If the child's initial grammar were "mixed", there is no number of inflected forms that could induce the child to reanalysis (Hyams 1991: 260).

We just saw how the null subject is *licensed* (i.e., what makes null subjects possible) by the child: they are licensed by morphological uniformity. The *identification* of the null subject (i.e., how the definite value of the null subject is recovered) occurs in a different way depending on the type of language. In richly inflected languages, such as Italian or Polish, the null subject is identified by AGR, both by children and adults. As children acquiring these languages learn the inflectional system early, it is reasonable to assume that they are able to identify the null subject by AGR. In poorly inflected languages, such as English, the null subject produced by children is identified by a topic, as for Chinese and other languages which uniformly lack inflectional morphology.

Hyams' proposal also predicts that the child might allow null subjects, but disallow null objects. Consider the structures below in (13a) and (13b), a modification of Huang (1984):

(13)a. TOPIC_i [topic_i [pro_i INFL VP]]

b. *TOPIC_i [topic_i [NP INFL [V pro_i]]

In (13a), the sentence topic ('topic') is linked to a discourse topic ('TOPIC') which identifies *pro*. In (13b), the identification of *pro* is not possible. According to Huang's Generalized Control Rule (1984), empty arguments are identified by the closest identifier, which for *pro* in object position is the subject in (13b) and not the topic. However, the coindexation of *pro* with the subject NP would violate Principle B of the Binding Theory (Chomsky 1981). In true radical pro-drop languages, variables occur in (13) instead of a null pronoun (*pro*), so no Principle B violation occurs between the subject and the object in (13b); instead, the null object can be identified by the topic. However, coindexation with the topic is impossible in (13b), since the null object is a pronoun and requires identification by the closest identifier. Hyams (1991) proposes that in the early grammar, the child does not have variables, only little *pro* (see also Roeper et al. 1984), which cannot be identified without violation of Principle B.

When a child acquiring a language like English develops the agreement system, she switches to agreement-identification of the null subject instead of topic-identification, which is reserved only to radical pro-drop languages. As only rich agreement systems can license the null subject, the English-speaking child realizes that null subjects are not possible in her language.

To sum up, according to Hyams (1991), English-speaking children go through two stages, which are represented in (14):

(14)a. *First stage*

Licensing of null subjects	*Identification of null subjects*
Children assume their language is morphologically uniform.	Children identify the null subject by a topic.

b. *Second Stage*

Identification of null subjects
Agreement wins over topic and null subjects are not identified anymore.

Italian-speaking children do not go through the two stages described in (14), since they correctly assume that their language is morphologically uniform (correctly licensing null subjects) and they quickly learn the agreement system of the language, which is responsible for identifying null subjects in their language (Italian-speaking children do not identify a subject by a topic). Children acquiring a radical pro-drop languages also do not go through the stages in (14): they also correctly assume that their language is morphologically uniform, licensing null subjects; as radical pro-drop languages lack agreement, they correctly identify null subjects by a topic.

In Hyams' (1991) analysis, the development of modals is still related to the child exiting the null subject stage, although in an indirect way. English modals require a [+tense] feature[53]. According to Hyams' (1991) approach, null subjects are not licensed anymore in English once the child acquires the inflectional system and realizes that her languages has a mixed paradigm. Once children realize verbs in their language have a [+tense] feature, they will produce modals.[54]

In Wang et al. (1991), the production of null subjects and null objects is compared between English- and Chinese-speaking children. The authors employed an elicited production task. In the earlier MLU stage tested (MLU = 3.5), Chinese- and English-speaking children do

---

[53] This is shown, for example, by the contrast below between (i), (ii) and (iii). A modal cannot appear after *to*:
(i) John hopes that Mary can come.
(ii) *John wants Mary to can come.
(iii) John wants Mary to come.

[54] There are dialects of English that allow modals in infinitives. In future research, it would be interesting to look at missing subjects in children exposed to these dialects (Bošković *p.c.* May 12*,* 2020). Under Hyams' (1991) account, the prediction is that, for children acquiring these dialects, the exiting of the null subject stage would not be related to the development of modals. In this respect, notice that sentences like 'I expect to can get a haircut next week' are possible in the South of Scotland (Gary Thoms, *p.c.* May 21, 2020).

not differ significantly in the use of null subjects, supporting Hyams' (1991) hypothesis that children start out with a radical pro-drop language. But they do differ in the production of null subjects at the later MLU stage (MLU = 4.5), suggesting that the English-speaking children's grammar has changed at this point. Furthermore, Wang et al.'s (1992) findings on the production of null objects are not predicted by Hyams (1991): Chinese-speaking children produce null objects along with null subjects (as adults), while English-speaking children only produce null subjects. In Hyams' (1991) analysis children acquiring Chinese are expected to not produce null objects, just like children acquiring English. Recall that in Hyams' analysis the reason why English-speaking children are not expected to produce null objects during the null subject phase is because they did not develop variables yet, so assuming they acquire variables at around the same time, English- and Chinese-speaking children should both reject null objects until they have developed variables. However, Wang et al. (1992) found that even after English-speaking children have developed variables (indicated by the fact that they produce wh-questions), they still produce null subjects but not null objects.

We have seen that Wang et al.'s (1992) data are problematic for Hyams' (1991) hypothesis that English-speaking children start out with a radical pro-drop grammar. If English-speaking children initially have a radical pro-drop grammar, then we would expect no difference in the early production of null objects in Chinese and English.

Missing subjects have also been analyzed as resulting from performance deficits (L. Bloom 1970, P. Bloom 1991, Valian 1990, Gerken 1991). Performance-based approaches claim that there is essentially no difference between the grammar of children and adults. That is, children drop subjects in the production of sentences because of their limited performance capacity (i.e., limited working memory). For example, P. Bloom (1991) proposes that children

omit subjects at the onset of a sentence in function of the processing load being greatest when the child starts producing a sentence. A study by Orfitelli and Hyams (2008) provides strong evidence against purely performance-base accounts. The authors show that English-speaking children who drop subjects also accept subjectless sentences.

In section 3.6, I return to the issue of missing subjects in English, providing an analysis that does not face the challenges described above.

1.2 Brazilian Portuguese

Simões (1999) conducted a longitudinal study on null subjects in a child acquiring Brazilian Portuguese as his native language: André, age 2;4 to 3;0. Simões' study focuses on the production of definite null subjects. In order to calculate the proportion of null subjects produced by André, the author excluded from the counting imperative sentences, answers to 'yes/no questions' (since they generally have a null subject in BP)[55] and nonfinite clauses. The table below shows that the use of null subjects by André remains relatively stable over time:

Table 1: Null subjects and overt subjects in André's speech

Age (record)	**Null Definite**	**Null Expletive**	**Personal Pronoun**	**Demonstrative Pronoun**	**Overt NP**
2;4.14 (A1)	54.9%	-	33.1%	-	12%
2;4.21 (A2)	39.7%	5.5%	31.9%	16.4%	6.5%

[55] An example to answer to a 'yes/no question' in BP is given below:

Q: O João foi pra escola?
The John went to school
'Did John go to school?'

A: *e* foi.
went
'He did.'

2;5.5 (A3)	52.0%	4.1%	28.4%	5.9%	10.6%
2;5.19 (A4)	58.5%	7.4%	19.1%	8.5%	6.3%
2;6.2 (A5)	32.1%	8.0%	31.2%	10.6%	18.1%
2;6.16 (A6)	40.6%	5.0%	30.6%	15.1%	8.7%
2;7.9 (A7)	52.7%	8.5%	18.6%	7.8%	12.4%
2;10.9 (A11)	51.5%	7.2%	27.7%	3.5%	10.1%
3;0.30 (A16)	34.0%	8.3%	43.2%	4.3%	10.2%
Average	46.2%	6.0%	29.6%	8.0%	10.5%

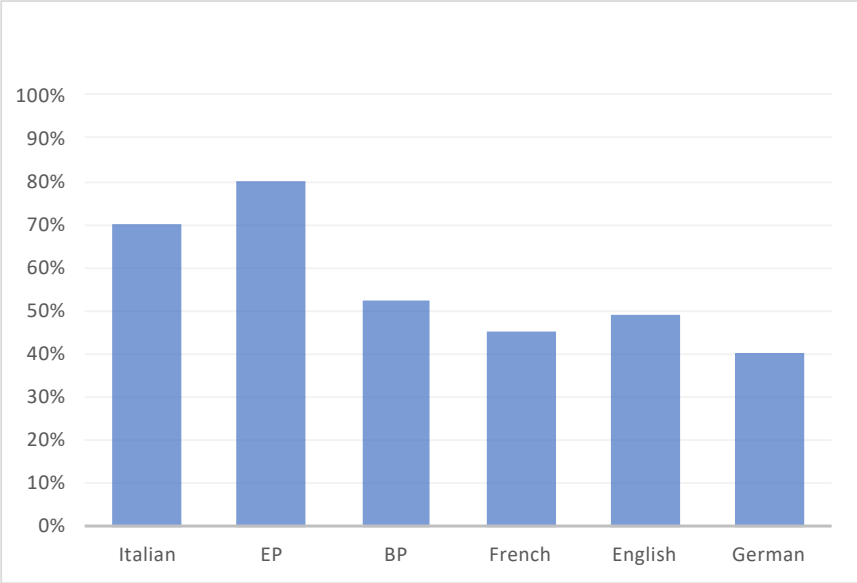
(Adapted from Simões 1999: 107)

Simões contrasts the use of null subjects by André and Adam, an English-speaking child (the data are from Hyams and Wexler 1993) in two periods. The periods contrasted correspond to 2;5 and 3;0 for Adam and 2;4 and 3;0 for André. While in Adam's speech there is a dropping in the production of null subjects between the two periods (from more than 50% to less than 30%), the same cannot be observed in André's speech, where there is little change in the percentage of null subjects over time.

Simões also observes that the percentage of null subjects used by André is very different from the percentage of null subjects used by children acquiring consistent null-subject languages such as European Portuguese (EP) and Italian. According to Simões, the average of null subjects produced by André, from the age of 2;4 to 3;0, is comparable to the average of missing subjects in the speech of children acquiring non-null-subject languages within the same age range (with the difference that there is dropping in the frequency of missing subjects for children acquiring non-null-subject languages, which was not found in André's speech). The graph below compares the percentage of null subjects by André and children acquiring five other languages (Italian

(Valian 1991), EP (Faria 1993), French (Pierce 1992), English (Hyams and Wexler 1993) and German (Clahsen 1989)).

Graph 1: Null subjects in the acquisition of different languages[56]



(Adapted from Simões 1999: 110)

Simões (1999) compared the percentage of definite null subjects in André's speech with the percentage of definite null subjects found in BP adults, as reported by Duarte (1995). Recall, from Chapter 2, that Duarte studied the speech of 13 adults from Rio de Janeiro. The speakers were divided into three groups, according to their ages: Group 1 (59- to 74-year-olds), Group 2 (45- to 53-year-olds) and Group 3 (25- to 32-year-olds). Simões seems to have compared André's null subjects with the null subjects produced by the youngest speakers in Duarte's study (Group 3). For Group 3, the highest percentage of null subjects was 29% (third person null subjects).

[56] The exact percentage for each bar is not provided by Simões (1999).

Adjusting the initial counting of André's null subjects to the criteria used by Duarte, which excludes non-pronominal subjects, a total of 55.5% null subjects was found in André's speech.

Although the percentage of null subjects produced by André is higher than the percentage of null subjects reported by Duarte (1995), Simões (1999) argues that the distribution of null subjects in the child's speech follows the adult pattern. Particularly, null subjects occur in embedded clauses when controlled by an argument in the higher clause, and in matrix clauses when the antecedent can be easily identified by the discourse (see Chapter 2). There are also cases in which the antecedent is absent in the discourse, but very salient in the pragmatic context, as the example in (15) shows. Simões points out that cases like (15) are extremely marked in BP, as the context must be immediate enough to make clear the reference of the null subject.

(15) [context: two friends are talking in the kitchen while cooking]

- Ih, *e* queimou!

Uh, burned

'The food burned!'

While contexts such as (15) can be argued to be not so common in the speech of adults, who are frequently talking about non-immediate situations (i.e., adults often talk about the past, future etc.), they are common in the speech of children. Children talk about the "here and now" and the fact that this is a common context for André explains, at least partly, his high percentage of null subjects compared to the speech of adults (Simões 1999).

According to Simões (1999), two other factors contributed to the high number of null subjects found in André's speech: what the author calls "ritualized repetitions" of a utterance (16)

and the frequency with which André answers a question by using the expression in (17), which is usually uttered with a null subject, even in the adult's speech (however, an adult is unlikely to keep answering a question as "I don't know" multiple times in a conversation).

(16) *e s*ai e *e* põe, *e* sai e *e* põe. (André 2;6)
get.out:3SG and put.in:3SG get.out:3SG and put.in:3SG
'It gets out, I put it in, it gets out, I put it in again.'

(17) *e* não sei.
not know:1SG
'I don't know.'

Simões concludes that there is no evidence of parameter missetting in the speech of André: all contexts in which the child uses null subjects are grammatical for BP-speaking adults. The somewhat high frequency of null subjects in André's speech can be explained by the child's use of null subjects in three contexts where null subjects are uncommon in the adult's speech, nevertheless, possible: (i) "here and now" contexts; (ii) the number of times in which André answers a question by using the utterance in (17) and (iii) "ritualized repetitions".

Like Simões (1999), Lopes (2003) presents a longitudinal study on the acquisition of null subjects by a single child acquiring BP, Raquel. The difference is that Raquel was studied for a longer period of time than André, starting from 1;9 to 3;0 (recall that André was studied from the age of 2;4 to 3;0). The author identified two different stages for the production of definite null subjects: in the first stage, from 1;9 to 2;8, the child produces 44.1% definite null subjects, which

is more than found in the adult's grammar. In the second stage, when the child reaches 3;0, the production of null subjects drops to 32.9%, a percentage that corresponds to adult levels.

Observe that the pattern of production of null subjects by Raquel resembles the one found in André's speech (Simões 1999). Although Raquel initially produces more definite null subjects than usually found in the adult's grammar, the percentage of null subjects in Raquel's speech does not drop drastically, contrarily to what is found in English. Also, Raquel produces fewer null subjects than children acquiring consistent null-subject languages, whose percentage of null subjects fluctuates between 70% and 80% (see Graph 1).

Magalhães (2006) conducted a longitudinal study on the development of subjects and objects in two children acquiring BP, Raquel (BP) (ages 1;9 to 3;0) and Ana (BP) (2;4 to 2;10) (Raquel is the same child studied by Lopes (2003)). Magalhães also studied, in the same work, the acquisition of subjects and objects by two Portuguese children, João (EP) (2;0 to 2;7) and Raquel (EP) (1;10 to 2;11) and compared the results of children acquiring the two languages. For our purposes, I will report only the results for subjects.

Magalhães (2006) includes in the counting definite null subjects in matrix clauses with finite verbs. She excluded imperative sentences, explicit cases of repetitions from sentences uttered by adults, sentences uttered in tag-contexts and answers to yes-no questions using 'ser' (*to be*). Differently from Simões (1999) and Lopes (2003), Magalhães (2006) included in the counting answers to yes-no questions other than with 'ser'.

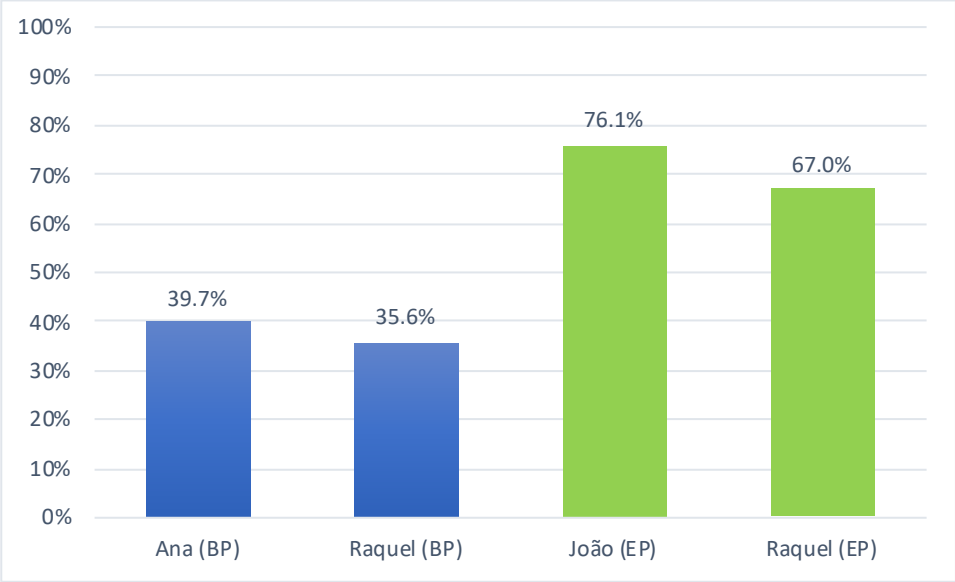
Magalhães identifies the same two stages previously identified by Lopes (2003) in the use of null subjects by Raquel (BP): the first stage goes from 1;9.8 to 2;8.14 and the second one happens in the last recording, when the child is 3;0.15. In the first phase, the average of definite null subjects is 52.75% and in the second phase, 35.6% definite null subjects were registered.

The author also identifies two stages in the speech of Ana (BP): the first one goes from 2;4.11 to 2;10.2, where there is an average of 69.81% definite null subjects and the second one happens, as in the case of Raquel (BP), in the last recording (2;10.29), when the percentage of null subjects in the child's speech approaches adult levels by dropping to 39.7%.[57]

For the Portuguese children, although there is some oscillation between sessions, the percentage of null subjects remains high from the first to the last session. João's (EP) average of definite null subjects is 71.21%. For Raquel (EP), the average of definite null subjects is 70.32%.

The graph below summarizes the different percentages of null subjects in the speech of Brazilian and Portuguese children. Percentages refer to the *last* session analyzed for each child:

Graph 2: Percentage of null subjects in Brazilian and Portuguese children



(Adapted from Magalhães 2006: 71)

[57] For both Raquel (BP) and Ana (BP), as the last stage includes only one recording, it is unknown whether the children would sustain a low percentage of null subjects if they were recorded for a longer period of time or not.

The definite null subjects that remain in the speech of Brazilian children happen in two main contexts, according to Magalhães: answers to questions (18) and null subjects that are identified by the physical presence of the referent (19). In the speech of Portuguese children, the null subject appears in various contexts, according to the author.

(18) MOT: que o homem (es)tá fazen(d)o? (Raquel (BP) 3;0)
what the man is doing
'What is the man doing?'
CHI: (es)tá catando.
be:3SG picking.up
'He is picking that up.'

(19) CHI: <ai@i (es)tá duro ## ajuda>[<]! (Raquel (BP) 2;8)
ah be:3SG hard help
'Ah, it is stuck, help!'
[trying to open a box]
(Magalhães 2006: 133-4)

According to Magalhães (2006), the definite null subject in (18) is identified by the topic of the sentence. This is a topic as described by Kiss (1995): an expression denoting an individual or a group already established in the discourse. For cases like (19), in which the referent is deictic, the physical presence of the referent in the context identifies the null subject. Magalhães

adopts Modesto's (2000b) proposal to explain cases like (18), in which the reference of the null subject is recovered by the topic. Consider the sentences below in (20):

(20)a. O João$_i$ disse que e_i comprou um carro.
The John said that bought:3SG a car
'John said that he bought a car.'

b. A Maria$_i$, o João disse que e_i comprou um carro.
The Mary, the John said that bought:3SG a car
'John said that Mary bought a car.'

c. Q: O que o Deputado Roberto Jefferson fez?
What that the Deputy Roberto Jefferson did:3SG
'What did Deputy Roberto Jefferson do?'

A: *e* denunciou o esquema do mensalão.
reported:3SG the scheme of.the mensalão
'He reported the *mensalão*.'

In Modesto's (2000b) approach, the empty category in (20a) is identified by the overt subject in A-position. The empty category in (20b) is a *pro* bound by the subject in an A'-position. In (20c), where we have an answer to a question, as in null subjects produced by children, the null subject is a trace of a deleted topic phrase.

In this section, I discussed the production of null subjects by children acquiring BP in three studies: Simões (1999), Lopes (2003) and Magalhães (2006). I summarize below the important conclusions that can be extracted from these studies for our purposes:

(i) There is no evidence that children acquiring BP misset the null subject parameter(s): all contexts in which the child uses null subjects are grammatical in the adult's speech.

(ii) Children acquiring BP use fewer null subjects than children acquiring consistent null-subject languages, such as EP.

(iii) Definite null subjects in matrix clauses in BP are of three types in the grammar of both adults and children acquiring BP: (a) they can be a trace of a deleted topic, (b) they can be null subjects recovered by the physical presence of the referent, or (c) they can be embedded null subjects in finite clauses controlled by an element in the higher clause.

2 Previous studies: Impersonals

In the last section, I reviewed studies on the acquisition of definite subjects by children acquiring English, BP and EP. In this section, I turn to the acquisition of impersonal structures. I start by discussing spontaneous production data by BP-speaking children, showing that generic null subjects appear before the age of 3;0. Next, I turn to the acquisition of impersonals in Estonian, a partial null-subject language, like BP and Finnish. In the last subsection, I review studies showing that children acquiring EP and Spanish also produce impersonal constructions early on, but with *se*, in accordance with the adult grammar.

2.1 Brazilian Portuguese

Simões' study, which was discussed in the previous section, focuses on the production of definite null subjects. The use of generic null subjects was not quantified by the author, however, she mentions that André also produces null subjects of this type. The author shows the following occurrences of generic null subjects in the data:

(21)a. Como é que *e* tila [*] [= tira]? (André 2;10)
How is that take.off:3SG
'How does one take this off?'

b. Como é que, como é que *e* anda? (André 3;0)
How is that, how is that walk:3SG
'How does one walk?'

c. Como é que *e* faz? (André 3;0)
How is that do:3SG
'How does one do it?'

d. Como é que *e* entra agora. (André 3;0)
How is that enter:3SG now
'Now, how does one get in?'

(Simões 1999: 126)

Although Simões does not focus on the acquisition of impersonal structures, the data in (21) are important for our purposes, as they show that a child acquiring BP can produce generic null subjects before the age of 4;0 (recall that in Experiment 1, discussed in Chapter 4, the youngest age group tested was 4-year-olds).

Magalhães (2006) also reports generic null subjects in the speech of Raquel and Ana, the two Brazilian children she investigated.[58] Like Simões, the author did not quantify the occurrences of generic null subjects.

(22)a. Com(o) é que *e* abre esse negocinho aqui? (Raquel 3;0)
How is that open:3SG this little.thing here
'How does one open this little thing here?'

b. aí # <como *e* faz isso> [/] como *e* faz isso? (Ana 2;10)
here how do:3SG this how do:3SG this
'How does one do this?'

(Magalhães 2006: 137)

Notice that in all examples above the child asks for instruction on "how to do something". Instructions are rules on how something should be done. In this sense, it can be said that the sentences in (22) and (23) have a lawlike background. It is also worth mentioning that the definite subject is unacceptable in sentences like (21) and (22), even when salient in the

[58] Two of Magalhães' examples (67b and 67d, pp. 137-8) were omitted here, since it is not clear whether the null pronouns in the examples have a generic interpretation or not.

pragmatic context or physically present. The definite null subject in (21) and (22) is impossible also when controlled by an overt subject in a higher clause.[59] Somehow, these sentences seem to license a generic operator, but it is not clear how the [+wh] element would function as an overt marker of genericity, if it does. This question is left for future research.

We saw that the use of null impersonals in BP with the generic interpretation was found very early in the speech of three children, with the earliest use registered in the speech of André and Ana at the age of 2;10.[60]

I will now turn to the discussion of impersonal constructions in Estonian, a partial null-subject language.

2.2 Estonian

Torn-Leesik and Vija (2012) present the first longitudinal study on the acquisition of Estonian impersonals. Estonian is a language very closely related to Finnish, both being Finnic languages, and it is reported to be a partial null-subject language, at least in its spoken form (Holmberg, 2017: 366). Torn-Leesik and Vija (2012) investigated the production of impersonals in the speech of one child, Andreas, age 1;7 to 3;1 (CHILDES database)[61]. The authors' objectives were the following: (i) determine the first occurrence of impersonal constructions in the child's everyday conversations; (ii) identify the typical context in which the impersonal

---

[59] Consider, for example, the sentence below, in which the coreference between *João* and the null subject is not possible. The null subject in (i) receives a generic reading, regardless of the presence of *João*.
(i) *O João$_i$ perguntou como é que *e*$_i$ faz isso.
'João asked how one does that.'

[60] It is unknown whether generic null subjects occur before 2;10 in the speech of children acquiring BP. A analysis of longitudinal data focusing on null impersonal constructions in BP needs to be conducted in the future. To my knowledge, there is no longitudinal study specifically addressing the acquisition of null impersonal constructions in BP.

[61] CHILDES (Child Language Data Exchange System) http://childes.psy.cmu.edu/

construction occurs in the child's speech; (iii) identify possible relations between impersonal constructions in the child's speech and his caregiver's speech.

The Estonian impersonal is often considered to be the basic voice construction in Estonian (Viitso 1998). The Estonian impersonal can be formed with either transitive or intransitive verbs, with unaccusatives and modals (Torn 2002; Torn-Leesik 2007, 2009). Notice in the examples below that the Estonian impersonal sentences have no overt subject, which makes this study interesting for our purposes, since the impersonal structures that we are studying in BP also have a null subject. In the present tense, the suffix *-(d/t)akse* on the verb (23) indicates that the voice in the construction is impersonal. In the past tense, the suffix *-d/ti* marks impersonality on the verb (24). In the present perfect, the impersonal is formed by the third person present form of the auxiliary *olema* ('to be') and the *-tud* participle marker on the main verb (25). The past perfect is formed similarly to the present perfect, the only difference being that the auxiliary *olema* is in the third person past form (26):

(23) Loetakse raamatuid.
read:IMP:PRES books:PART
'One reads books.'

(24) Loeti raamatuid.
read:IMP:PAST books:PART
'One read books.'

(25) On loetud raamatuid.

be:PRES:3 read:PASS:PTC books:PART

'There has been read some books.'

(26) Oli loetud raamatuid.

be:PAST:3 read:PASS:PTC books:PART

'There had been read some books.'

(Torn-Leesik and Vija 2012: 252)

The interpretation of the impersonal subject in Estonian can be generic (equivalent to *one*), but as the authors describe, it can also refer "to an indefinite actor whose identity is unknown or is left unspecified for considerations of relevance or politeness." Additionally, the impersonal can be used for "for salient or identifiable referents" for reasons of politeness again or "to express stylistic nuances." (Torn-Leesik and Vija 2012: 553).

Studies on the comprehension of impersonal structures in Estonian show that children comprehend these structures by the age of four (Vija et al. 2009 and Kunnari et al. 2010). Similarly, in Experiment 1 (Chapter 4), I show that children acquiring BP at the age of four already understand that the null subject in impersonal constructions has a non-definite interpretation.

Each of Andreas' recording sessions lasted approximately 1 hour. He started being recorded when he was 1;7.24 until the age of 3;1.13. The authors divided the recordings into four different age periods: 1;7-2;0, 2;0-2;1, 2;3-3;0 and 3;0-3;1. The first impersonal forms appeared in Andreas' speech in the second period (2;0-2;1), but at this point these forms seem to be imitations from the input and they are not completely adult like (the forms are truncated and,

mainly, they do not fit into a coherent conversation, suggesting that the child does not know the contexts in which impersonal forms are used). The authors suggest that in the second period impersonals were learned as "chunks". However, during this period they also report correct use of impersonals which does not seem to be a direct repetition from the input: at the age of 2;0.23 Andreas' father tells him to not pour milk into the sink, without using an impersonal structure (the negative indicative is used instead), and Andreas replies using an impersonal structure *kallatakse* ('one pours'):

(27) FAT: ei kalla kraanikaussi piima.
'no pouring of milk into the sink!'
CHI: piima **kallatakse**.
milk:PART pour:IMP:PRES
'one pours milk.'
FAT: piima **kallatakse** suhu ja snust läheb kõhtu.
'one pours milk into the mouth and from the mouth it goes to the stomach.'

(Torn-Leesik and Vija 2012: 259)

From the third period on, Andreas' use of impersonal forms advances to a large extent, according to the authors. Although there are problems sometimes in the choice of the stem, the impersonal suffix is used correctly. The use of impersonals is also coherent with the context in the last two periods investigated.

Most of the impersonals used by Andreas are in the present tense affirmative. After the age of three, instances of past impersonals appear in his speech, although they remain rare in his production (the authors reported that past impersonals are also rare in the input of the parents). The authors also found that apparently there is a correlation between the number of different impersonal verbs used by Andreas and the number of impersonal verbs found in Andreas' mother (most of the conversations happen with his mother, but a few of them with his father). In the second period (2;0 - 2;1.12) Andreas' data show 13 impersonal forms (tokens) of 7 different verbs (types), while in the fourth period (3;0 - 3;1.13), there are 139 impersonal forms (tokens) of 38 different verbs (types). The number of different verbs in impersonal constructions used by Andreas' mother is reported to increase as well during this same period: the tokens increase from 53 to 175 and the types from 26 to 56. Although the authors do not provide an interpretation for this fact, it could be just an indication that the verbal interaction between mother and child becomes more complex and richer as the child gets older, with both mother and child using more linguistic forms to interact with each other: this would imply that not only the use of impersonal forms increase, but also the use of other grammatical constructions, a fact that was not investigated by the authors.

As for the contexts in which impersonal forms are used by Andreas and also by his parents, the authors report that they are of a general character. They describe how people usually behave, they state conventional rules or habits, as in the example in (28). Negative impersonals used by the parents mostly express rules and prohibitions. For our purposes, it is crucial to notice that this use of impersonal constructions is *generic*. The fact that the child uses them productively already by the age of 2;3 provides us with an indication that genericity is mastered very early on. If we could use a simpler methodology than the TVJT used in Experiment 1

(Chapter 4), it is possible that evidence from children's comprehension of impersonal structures would be found before the age of 4.

(28) MOT: me lähme bussiga sinna kopsuartile.
'we're taking the bus to the lung doctor.'

CHI: aga kas kaheaasta [*] [= kaheaastasele] **tehakse** süsti?
but Q two-years old:ALL do:IMP:PRES injection:PART
'but are shots given to two-year-olds?'

MOT: kui lapsed kaheseks saavad, siis tehakse jah
when children two:TERM become:PRES.3PL then do:IMP.PRES yes
kaitsesüsti, vaktsineerimist.
protective-injection:PART inoculation:PART
'when children turn two, then they are indeed given a protective shot, an inoculation.'

(Torn-Leesik and Vija, 2012: 264)

I summarize below the important conclusion that can be extracted from this study for our purposes:

(i) Use of null impersonals in Estonian (a partial null-subject language) with the generic interpretation was found very early in the speech of one child. Consistent adult-like use of the construction started at the age of 2;3.

2.3 Spanish

Seabrooks (2017) investigated the production of transitive, passive and impersonal constructions in the spontaneous speech of four Spanish-speaking children: Eduardo, Emilio, Maria and Iago. The age range of the children was 1;4 - 4;6. As impersonals in Spanish have the clitic *se*, it is helpful to compare the acquisition of impersonals constructions with the clitic *se* in Spanish and in BP.

The objective of the author was to test Mendikoetxea's (2008) hypothesis that passives and impersonal constructions are distinct structures. Considering the examples in (29) and (30), some theories have suggested that *libros* ('books') is a syntactic subject in both impersonals and passives (Suñer 1976). In this perspective, the constructions are structurally identical with the difference located only in the subject-verb agreement pattern. As can be seen in (29), the verb in the impersonal structure does not agree with the object *libros* ('books'), but in the passive in (30), the verb agrees with the logical object *libros*. In BP, the structure of *se* impersonals and passives with *se* is analogous to Spanish, as can be seen in (31) and (32), respectively:

(29) Aquí se vende unos libros. *Impersonal*
Here SE sell:3SG a.few books
'One sells a few books here.'

(30) Aquí se vende**n** los libros. *Passive*
Here SE sell:3PL the books
'Here books are sold.'

(Seabrooks 2017: 4)

(31) Aqui se vende livros. *Impersonal*
Here SE sell:3SG books
'One sells a few books here.'

(32) Aqui se vende**m** livros. *Passive*
Here SE sell:3PL books
'Here books are sold.'

Impersonal constructions like (29) are compared to simple transitive sentences like (33) in theoretical approaches that assume that *se* occupies the subject position (Oesterreicher 1992 and Rivero 2001). In these approaches, *libros* in (29) is the complement of *vender* (to sell), as in (33), and *se* in (29) is in the subject position, in a parallel way to *Juan* in (33):

(33) Juan vende los libros. *Spanish*
Juan sell:3SG the books
'Juan sells the books.'

(Seabrooks 2017: 4)

In her analysis of Romance clitic *se*/*si*, Mendikoetxea (2008) argues that that both impersonal *se* constructions and passives with *se* contain a generic null pronoun in [Spec, *v*P]. The difference between the two constructions is determined by case marking: the impersonal in

(29) has an active *v,* according to Mendikoetxea, which assigns accusative case to *libros*. On the other hand, in the passive construction, *v* does not case mark *libros*. The author also assumes that the passive and the impersonal construction are different from simple transitive sentences like (33) because while the impersonal and the passive construction have a generic null pronoun in [Spec, *v*P] (GNP), the transitive sentence does not. Mendikoetxea (2008) assumes the following structures for transitive sentences (34), impersonals with *se* (35), and passives with *se* (36). Notice that for Mendikoetxea (2008), while the impersonal in (35) has a null expletive, the passive (36) does not have it:

(34)a. [TP T [*v*P Juan [v' *v* [VP [V vende] [DP los libros]]]]]

b. [TP [DP Juan$_i$] [T' [T -e [*v*P t$_i$ [v' *v* [VP [V vend-] [DP los libros]]]]]

(35)a. [CP [NP Aquí] [C' **se** [TP [T' T [*v*P [DP GNP] [v' *v* [VP [V vende] [DP los libros]]]]]]]

b. [CP [NP Aquí] [C' **se** [TP 0 EXPL [T' T [*v*P [DP GNP] [v' *v* [VP [V vende] [DP los libros]]]]]]]

(36)a. [CP [NP Aquí] [C' **se** [TP [T T [*v*P [DP GNP] [v' *v* [VP [V venden] [DP los libros]]]]]]]

b. [CP [NP Aquí] [C' **se** [TP los libros$_i$ [T T -en [*v*P [DP GNP] [v' *v* [VP [V vend-] [DP t$_i$]]]]]]]

As passives and simple transitive sentences are beyond the scope of this thesis, I have no position on Mendikoetxea's analysis about the similarities and differences between simple transitive structures, passives and impersonal constructions. Neither am I assuming her approach to impersonals in this thesis (the approach that I am assuming is based on Holmberg (2005,

2010a,b), as was described in Chapter 2). I merely described her theory here, because Seabrooks' (2017) acquisition study aimed to test her hypothesis. Seabrook predicts that if the three structures (impersonals, passives and transitive constructions) are different from each other, children should not acquire them all at once, but at different points in time. Seabrooks also assumes that the order of acquisition parallels cross-linguistic frequency (Jakobson 1990). Taking into account that, cross-linguistically, actives are more frequent than passives, which in turn are more frequent than impersonals (Blevins 2003), the author predicts that the order of acquisition will be as following: *active —> passive —> impersonal*.

Regarding the acquisition of the clitic *se* in Spanish, Jackson-Maldonado et al. (1998) report that children produce the clitic *se* around the age of 2;4 to 3;0. Therefore, the acquisition of impersonal structures and passives with *se* is not predicted to occur before this age.

Considering all four children, 2964 simple transitive sentences were found (86.02%), 352 passives with *se* (10.21%) and 130 impersonals with *se (*3.77%). The numbers are given in tokens. According to Seabrooks, there is a common trend among children to acquire the constructions in the following order: simple transitive sentences —> passive expressions —> impersonal expressions, as predicted by the theory. But that is not quite the case: Maria acquired the simple transitive *with* the passive and Eduardo acquired the passive *before* the simple transitive. So, the pattern of acquisition that the author claims to have found happens with 2 out of 4 children. He is correct, though, in assuming that the impersonal structure appears later in the data.

For our purposes, the age in which the impersonal with *se* appears is relevant, in order to contrast it with the acquisition of impersonals with *se* in BP: according to Seabrooks, impersonals are used consistently only after 2;0. There is one instance of an impersonal in

Emilio's speech at the age of 1;8, but it appears to be a coding error, according to the author. However, in Maria's speech, the earliest tokens of impersonal constructions are found at the age of 1;11 (a total of 8 tokens). For Iago, the first impersonal construction is found at the age of 3;0. For Eduardo, the first impersonal is found at the age of 3;10.

As we saw, the acquisition of impersonals with *se* happens later in Spanish than the acquisition of passives and simple transitive sentences. Seabrooks provides two possible explanations for the later acquisition of impersonals with *se*. The first one depends on the fact that in impersonal structures, a phonologically null expletive merges with T, which does not happen in passives, according to Mendioketxea (2008). According to Seabrooks, the syntactic operation involving merging of the phonologically null expletive with TP (in impersonals) could be more challenging than merging the verbal complement with TP (in passives) and also more challenging than merging an overt subject. It is unclear why the insertion of the phonologically null expletive would be "more challenging" than the other operations.

Another explanation that the author provides for the later acquisition of impersonals with *se* in Spanish has to do with the generic null pronoun (GNP). For Mendikoetxea (2008), this pronoun appears in both impersonal constructions and passives, as can be seen in (35) and (36). Although the pronoun lacks the person feature, it has the number feature. Notice that while passives (36) have the plural marking on the verb, impersonals have a singular number feature. According to Seabrooks, this distinction between impersonals and passives with *se* can explain why children acquire passives before impersonals. He bases his argument on the fact that plural exclusive pronouns, such as *they*, are less marked than singular generic expressions, such as *one*

(Gelman et al. 2008)[62]. Impersonals would emerge later in the acquisition process because children are sensitive to this distributional constraint.

Below, I summarize the important conclusion that can be extracted from this study for our purpose:

(i) All children acquiring Spanish in the study productively use impersonal constructions with *se* before the age of 4;0.[63]

This is different from Brazilian children who show evidence for comprehension of impersonals with *se* much later than 4;0 as seen in Experiment 2 (Chapter 5). As discussed in Chapter 5, children younger than 6;0 behaved statistically worse than adults when the expected answer to the test sentence was *se*. Impersonals with *se* appear earlier in Spanish than in BP probably because impersonal with *se* is not very productive in BP anymore (Assis 2017).

2.4 European Portuguese

The focus of Magalhães' (2006) study, as mentioned in section 1.2, was the acquisition of definite null subjects by children acquiring BP and EP. However, the author reports occurrences of impersonal constructions with the clitic *se* in the speech of children acquiring EP. These

[62] Gelman et al. (2008) studied the acquisition of generic NPs in English. They found that although children talk about kinds early in development, indefinite **singular** generics are more restricted than mass nouns and bare plurals. Blevins (2003) points out that impersonal constructions with singular generic expressions are also less common cross-linguistically than impersonals with plural "generic" expressions (i.e., exclusive pronouns).

[63] Guasti (2002: 127) provides an example in which Diana, a child acquiring Italian, produces a sentence with impersonal *si* at the age of two.

(i) Si va al mare.
SI goes to.the beach
'We go to the beach.'

occurrences were not quantified by the author, but as can be seen in the examples below, they occur as early as 2;4 in European Portuguese (however, we do not know how productive the use of impersonal with *se* is by this age). Brazilian children, in contrast, are not reported to use impersonal with *se* in Magalhães' study (2006):

(37)a. co(l)a- se! (João 2;4)
glue:3SG SE
'One glues it.'

b. Agora põe- se. (Raquel 2;7)
now put:3SG SE
'Now one puts it on.'

c. Abre- se como? (Raquel 2;8)
Open:3SG SE how
'How does one open it?'

3 Setting the Null Subject Parameters

In this section, I propose an analysis of how children set the null subject parameters in their language. This analysis is based on the hierarchical organization proposed by Holmberg (2010a) (see (38a) and (38b)). The proposal aims to include and explain the acquisition of a complex null-subject typology, including radical pro-drop languages, consistent null-subject languages, partial null-subject languages and semi null-subject languages.

3.1 D in T, P in T and φ-dependent EPP

Holmberg (2010a) proposes three parameters to account for the distribution of null subjects: D in T, P in T and ϕ-dependent EPP. The first two parameters, as we will see in detail below, are related to each other (in a "bleeding relation", as the author describes it), while ϕ-dependent EPP is an independent parameter.

According to Holmberg (2010a), in languages with a positive value for ϕ-dependent EPP the only category that can satisfy the EPP is the goal of T's probing, that is, the DP or pronoun which values T's uϕ features and is assigned nominative. When a language has a negative value for the ϕ-dependent parameter, the EPP can attract other categories, such as adverbials.

As observed in Chapter 2, in this thesis I use the term EPP pretheoretically: I do not assume that there is an EPP feature in T which needs to be checked. Therefore, the ϕ-dependent parameter is reinterpreted as follows: in languages with a positive setting for the ϕ-dependent parameter, [Spec, TP] is filled by the DP or pronoun which values T's ϕ-features[64]. In languages with a negative value for the ϕ-dependent parameter, [Spec, TP] can be filled by an element which does not value T's ϕ-features.

The D in T parameter states that there is a definiteness feature (D-feature) which gives the definite interpretation to a null element. Consistent null-subject languages are the ones with a positive value for D in T, generating null subjects with definite reference. Languages with a negative value for D in T do not allow null subjects with definite reference: these include partial null-subject languages, semi null-subject languages and non-null-subject languages.

[64] See Bošković (2007) for a way of stating this in terms of case, where, essentially, nominative case in such languages can only be licensed in [Spec, TP].

Polysynthetic languages have D in all heads (C, T and *v*), generating null pronouns with a definite interpretation in any argumental position.

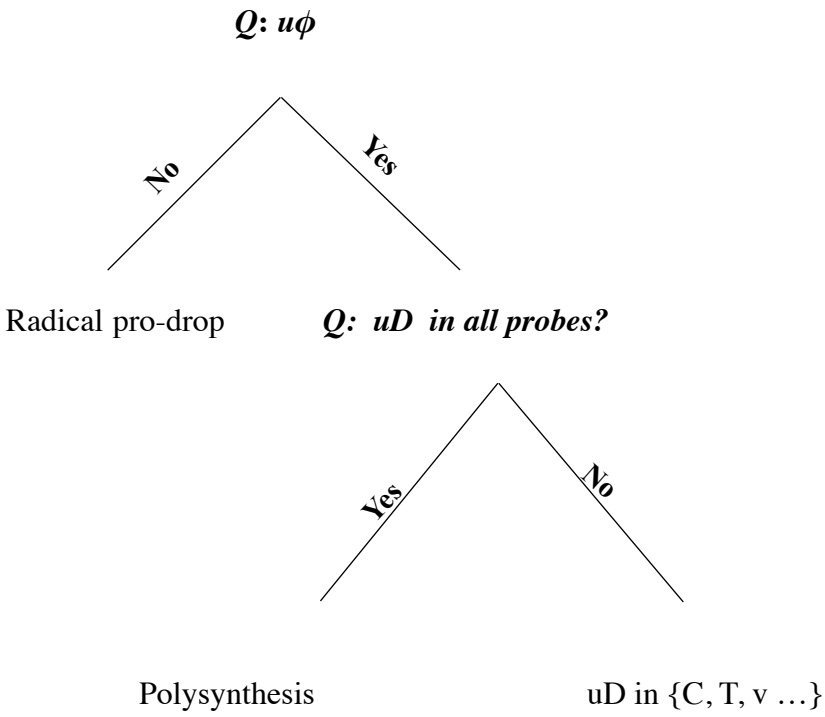
A language with a positive value for the P in T parameter forces the pronunciation of the subject in finite [Spec, TP]. The positive value of this parameter accounts for languages like Icelandic which do not allow control of a null subject in finite clauses (since the pronunciation of [Spec, TP] is forced), despite the fact that it allows generic null subjects. Icelandic allows generic null subjects because they are not in [Spec, TP], but in [Spec, *v*P]. The negative value of the P in T parameter also accounts for languages that allow expletive null subjects, such as Kriyol (Guinea-Bissau Creole), but do not allow definite null subjects: a null [Spec, TP] is therefore possible in this language.

The D in T parameter and the P in T parameter are in bleeding relation in the sense that a positive setting for the D in T parameter is necessarily a negative setting for the P in T parameter, and a positive setting for the P in T parameter is necessarily a negative setting for the D in T parameter (Holmberg 2010: 121, footnote 29). However, as some languages have the negative setting for both parameters, a negative value for the D in T parameter does not necessarily render a positive value for the P in T parameter.

Additionally to these three parameters, the hierarchical organization includes the presence or absence of uϕ-features in heads. Radical pro-drop languages have a negative value for uϕ, since they lack agreement, while all the other languages have a positive value for it. It is the lack of uϕ-features in the grammar which is responsible for null arguments in radical pro-drop languages (Roberts and Holmberg 2010). In (38a) we see this first split between radical pro-drop languages, which do not have uϕ-features in their grammar, and all the other languages, which have those features in their grammar. The second split in (38a) is between languages with uD in

all heads (polysynthetic languages such as Mohawk)[65] and languages which have uD in some heads[66].

(38)a.



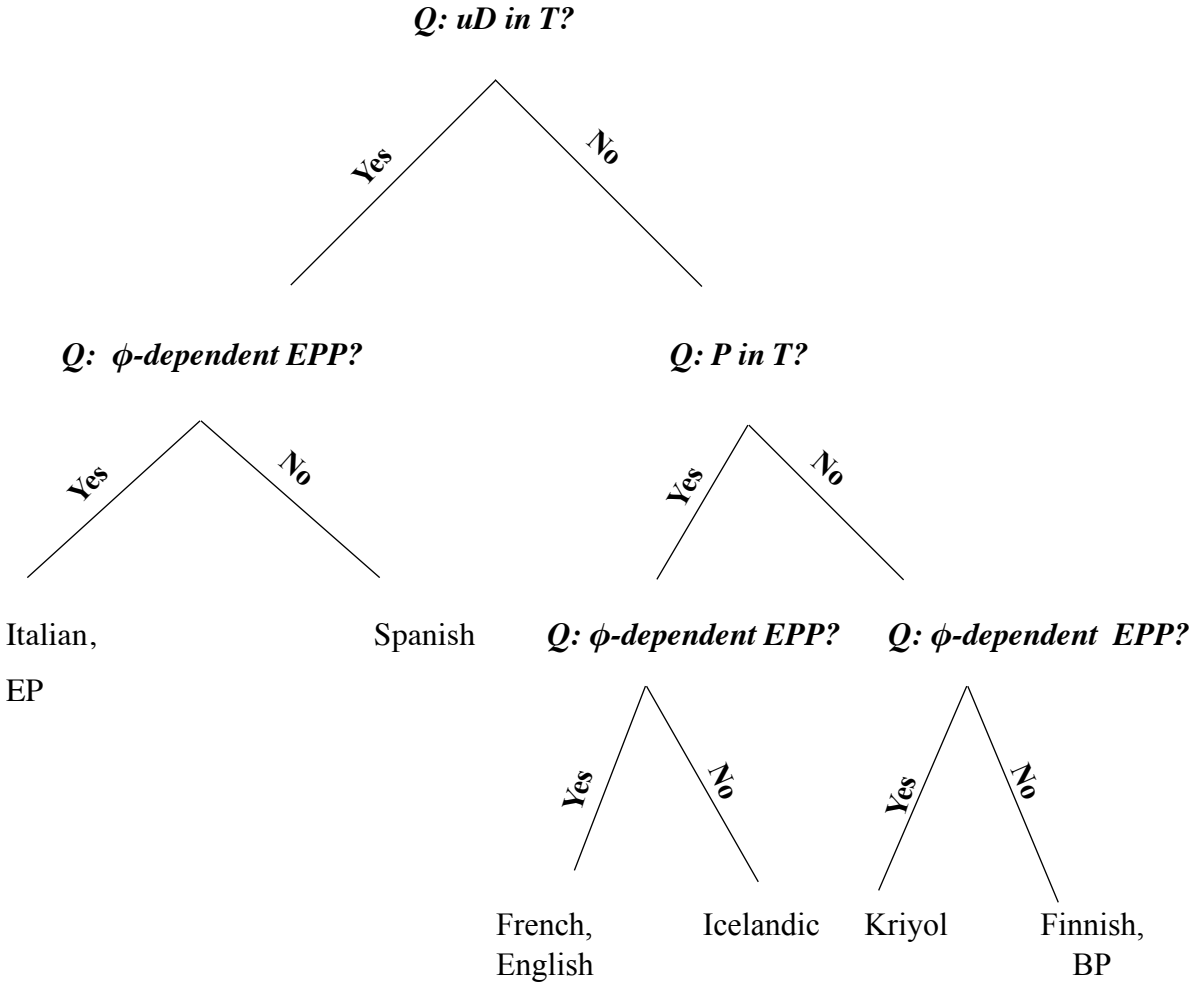
(adapted from Holmberg 2010a: 122)

[65] Presumably, the presence of D(efiniteness)-feature in a language is not associated with the presence of articles in this account, as all polysynthetic languages lack articles. Also, there are consistent null subject languages such as Serbo-Croatian that lack articles even though consistent null-subject languages have a D-feature in T.

[66] German is an example of language with a D-feature on C, given the fact that German subordinate clauses introduced by a complementizer have empty arguments in [Spec, CP]. See Berger-Morales and Salustri (2001) for details of the analysis. Pashto is an example of language with a D-feature on *v,* which licenses consistent null objects (Roberts 2010: 78-9).

As we are interested in null subjects, only uD in T concerns us. The following is the hierarchical organization with the split between languages with uD in T and languages without uD in T:

b.



The table below gives the value of the parameter for each one of the languages in (38b).

Table 2: Null subject parameters across languages

Language	***uφ***	***D in T***	***P in T***	***φ-dependent***
Italian	+	+	-	+
European Portuguese	+	+	-	+
Spanish	+	+	-	-
English	+	-	+	+
French	+	-	+	+
Icelandic	+	-	+	-
Kriyol	+	-	-	+
Finnish	+	-	-	-
Brazilian Portuguese	+	-	-	-

As mentioned above, languages with uD in T are consistent null-subject languages, such as Italian, EP and Spanish. Notice that consistent null-subject languages can either have a positive value for ϕ-dependent EPP, which renders Italian and EP, or a negative value for the ϕ-dependent EPP, which renders Spanish. As we see in the example below in (39), Spanish can have an indirect object (PP) in [Spec, TP], while T agrees with 'los chocolates'. This fact makes Spanish a ϕ-independent language. An analogous sentence to (39) is not possible in Italian or EP, which makes them ϕ-dependent languages.

(39) A Juan le gustan los chocolates. *Spanish*
to Juan CL like:3PL the chocolates
'Juan likes chocolates.'

(Holmberg, 2010a: 112)

A positive value for the D in T parameter gives necessarily a negative value for the P in T parameter, since the presence of definite null subjects means that [Spec, TP] does not need to have an overt (pronounced) subject. All other types of null-subject languages (partial, non-null-subject and semi null-subject) do not have D in T.

The next split in (38b) is the P in T parameter, which says whether the subject should be pronounced in finite [Spec, TP] or not. Non-null-subject languages such as English and French have a positive value for this parameter, as well as semi pro-drop languages, such as Icelandic, which need to have an element in [Spec, TP]. Notice that in the sentence below with a generic null subject in Icelandic, an adverbial appears in [Spec, TP] and the generic null subject in [Spec, *v*P]:

(40) Nú má *e* fara að dansa. *Icelandic*

Now may one go to dance

'One may begin to dance now'

(Sigurðsson and Egerland 2009: 169)

Partial null-subject languages, which always allow null subjects in subordinate clauses, and semi pro-drop languages such as Kriyol, which allow only null expletives, have the negative value for the P in T parameter.

As can be seen in (38b), English and French have ϕ-dependent EPP, since the EPP necessarily attracts an overt subject assigned nominative case, while Icelandic has ϕ-independent

EPP. In Icelandic, the EPP does not necessarily attract the goal of T's probing, but the closest category to T. This can be seen in (41):

(41) Mér voru gefnar peninga. *Icelandic*
Me were given money:PL
'I was given money.'

(Holmberg 2010a: 107)

In (41), the EPP attracts the highest DP ('mér'), which has inherent dative case[67]. The goal of T's probing is the direct object 'peninga', which is assigned nominative case. As the EPP attracts the closest category to T ('mér'), despite the fact that 'peninga' is the goal of T's probing, Icelandic has ϕ-independent EPP.

Regarding languages with the negative value for the parameter P in T, Kriyol has ϕ-dependent EPP, while Finnish and BP have ϕ-independent EPP. In Finnish and in BP, as discussed in detail in Chapter 2, an element other than the goal of T's probing, such as adverbials, can appear in [Spec, TP]. On the other hand, Kriyol has the positive value for the ϕ-dependent EPP parameter: elements that are not the goal of T's probing cannot appear in [Spec, TP].

3.2 Morphological Pattern and Subject Drop

[67] Quirky subjects like *mér* may actually also have a structural nominative case, which is not morphologically realized. This is so because *mér* can only be located in cased [Spec, TP] and cannot move away from a cased [Spec, TP] (Boeckx 2003, Bošković 2007).

Before discussing how children acquiring different languages set values for uϕ, ϕ-dependent EPP, D in T and P in T, I shall discuss to what extent overt morphological patterns are related to the possibility of subject drop. As is shown below, although it seems to be true that consistent null-subject languages necessarily have rich agreement, richness of agreement does not ensure that a language will allow pro-drop (as in the case of Finnish). Neither is it true that languages that completely lack agreement on the verb will allow pro-drop (as in the case of Kriyol). In other words, the hypothesis that morphological uniformity licenses null subjects put forward by Jaeggli and Safir (1989) does not seem to hold. Consider the inflectional paradigm of the languages below (these are the same languages from Table 2):

(42) **Radical pro-drop**

Chinese

'tánlùn' (talk)

1SG	tánlùn	1PL	tánlùn
2SG	tánlùn	2PL	tánlùn
3SG	tánlùn	3PL	tánlùn

No distinction

(43) **Consistent null-subject languages**

Italian

'parlare' (talk)

1SG	parl**o**	1PL	parl**iamo**
2SG	parl**i**	2PL	parl**ate**

3SG parl**a** 3PL parl**ano**

6 distinctions

European Portuguese

'falar' (talk)

1SG	fal**o**	1PL	fala**mos**	
2SG	fala**s**	2PL	fala**m**	(vocês)[68]
3SG	fala	3PL	fala**m**	

5 distinctions

Spanish

'hablo' (talk)

1SG	habl**o**	1PL	habl**amos**	
2SG	habl**as**	2PL	habl**áis**	
3SG	habl**a**	3PL	habl**an**	

6 distinctions

(44) **Non-null-subject languages**

English

'talk'

1SG	talk	1PL	talk	
2SG	talk	2PL	talk	

[68] The second person plural pronoun *vós* with the verbal inflection **-is** (fala**is**) is hardly used in spoken language. *Vocês* is used in spoken language instead of *vós* with the verbal inflection **-m** (fala**m**).

3SG talk**s** 3PL talk

2 distinctions

French[69]

'parler' (talk)

1SG paʁl 1PL paʁl**ɔ̃**

2SG paʁl 2PL paʁl**e**

3SG paʁl 3PL paʁl

3 distinctions

(45) **Semi null-subject languages**

Icelandic

'tala' (talk)

1SG tala 1PL töl**um**

2SG tala**r** 2PL tal**ið**

3SG tala**r** 3PL tala

4 distinctions

Kriyol

'fala' (talk)

[69] The phonological transcription was given for French, since the inflections seen in the written language (i.e., parl**e**, parl**es**, parl**e**, parlo**ns**, parle**z**, parl**ent)** are different from what is heard in the spoken language. While five distinctions are seen in the written form in French, only three distinctions are heard in the spoken language.

1SG	fala	1PL	fala
2SG	fala	2PL	fala
3SG	fala	3PL	fala

No distinction

(46) **Partial null-subject languages**

Finnish: Standard colloquial dialect

'puhua' (talk)

1SG	puhu**n**	1PL	puhu**taan**
2SG	puhu**t**	2PL	puhu**tte**
3SG	puhu**u**	3PL	puhu**u**

5 distinctions

Brazilian Portuguese

'falar' (talk)

1SG	fal**o**	1PL	fala
2SG	fala	2PL	fala**m**
3SG	fala	3PL	fala**m**

3 distinctions

Notice that both Chinese and Kriyol do not have person and number distinctions on the verb.[70] While Chinese is a radical pro-drop language, Kriyol does not license thematic null subjects (only null expletives are possible in the language). In other words, it is not true that all languages without any complex form in the verbal paradigm necessarily license null subjects, although it seems to be usually the case. Also, comparing Finnish and European Portuguese, we see that both languages make five distinctions in the verbal paradigm. However, Finnish only licenses null subjects in certain contexts (it is a partial null-subject language), while European Portuguese licenses null subjects more generally (it is a consistent null-subject language). Therefore, the data above do not corroborate Jaeggli and Safir's (1989) theory that morphological uniformity licenses null subjects. The only generalization that can be put forward concerning morphological pattern is that *consistent null-subject languages have rich agreement,* but rich agreement or lack of agreement does not assure that a language will allow null subjects. Likewise, *radical pro-drop languages lack agreement,* but lack of agreement does not assure that a language will allow null subjects. In other words, we are dealing with one-way correlations.

3.3 Setting the uφ Parameter

Regarding the uϕ parameter, the child will assume that the language has a negative value for uϕ if the language does not have overt marking of person and number on the verb. If the child sees any overt marking of person and number on the verb, she will assume that the language has a positive value for uϕ. In this case, even minimal agreement marking on the verb, as we see in English, will be enough to tell the child that T has ϕ-features. Since children are

[70] Kriyol is said to have uϕ-features, even though the language does not overly mark person and number on the verb. This is possible using Chomsky's (2001) approach in which uϕ-features are purely abstract (Roberts *p.c.*, February 9, 2020)

highly sensitive to inflectional morphology (Slobin 1982; Wexler 1998; Guasti 2002), it seems reasonable that the child will look at the morphological pattern of the language to set the value of uϕ.

For the special case of Kriyol, which is said to have uϕ-features despite the fact that the language lacks overt marking of person and number (see note 70), the child will presumably assume that she is acquiring a radical pro-drop language, until the system tells her otherwise (e.g., lack of thematic null subjects and null objects). See section 3.6 for a detailed discussion of Kriyol.

In what follows, I will discuss the setting of the D in T, P in T and ϕ-dependent parameters for each language type (i.e., consistent null-subject languages, partial null-subject languages, semi null-subject languages and non-null-subject languages).

3.4 Consistent Null-subject languages: D in T, P in T and φ-dependent Parameters

Ideally, the child should be able to set the appropriate values for the D in T, P in T and ϕ-dependent parameters based only on positive evidence.

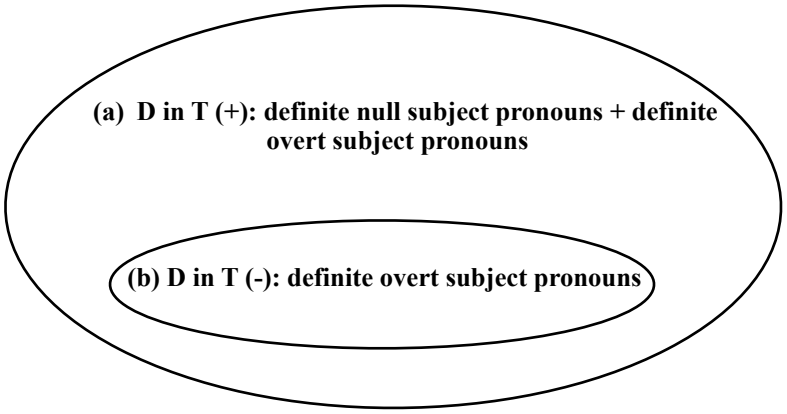
Let us first discuss the setting of the D in T parameter in consistent null-subject languages, such as Italian, EP and Spanish. These languages have a positive value for the D in T parameter and, consequently, a negative value for the P in T parameter (i.e., [Spec, TP] can be phonologically empty in these languages).

As discussed in section 3.2, rich agreement is not necessarily associated with pro-drop, as colloquial Finnish, a partial null-subject language, has the same number of distinctions in the morphological paradigm as European Portuguese (EP), a consistent null-subject language.

Consequently, rich agreement cannot be used as evidence to set a positive value for the D in T parameter.

The most direct evidence in the input to set the D in T parameter to a positive value is the presence of definite null subject pronouns under unrestricted conditions in matrix clauses. That is, as only consistent null-subject languages allow definite null subjects which are not recovered by a salient topic or the physical presence of the referent (recall that definite null subjects are possible in these contexts in BP, as discussed in section 1.2), these null subjects could be used as evidence by the child to set the positive value for the D in T parameter. Considering this evidence, the child would start out with a negative value for the D in T parameter (the subset in Figure 2) and later switch to the positive value for the D in T parameter (the superset in Figure 2). In other words, based on the Subset Principle, the negative value for the D in T parameter is the default value. However, as children acquiring consistent null-subject languages produce the same percentage of definite null subjects as adults, the setting of the D in T parameter apparently occurs very early in the child's development or it happens immediately.

Figure 2: D in T parameter and the Subset Principle

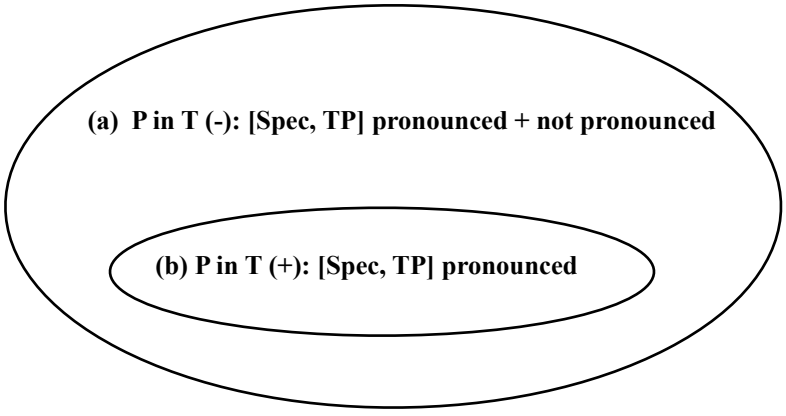


The default value for the P in T parameter is presumably positive, since languages that do not require the pronunciation of [Spec, TP] can have sentences in which [Spec, TP] is pronounced too, as seen by the pair below in Italian (47a,b). That is to say, the positive value for the P in T parameter is the subset and the negative value is the superset (Figure 3). The child would start out with the subset value, in which P in T is positive.

(47)a. Gianni va al mare. *Italian*
Gianni go:3SG to.the beach
'Gianni is going to the beach.'

b. Va al mare?
go:3SG to.the beach
'Is s/he going to the beach?'

Figure 3: P in T parameter and the Subset Principle



Once the child acquiring a consistent null subject language sets the positive value for the D in T parameter, based on the presence of non-restricted definite null subjects in the language, the negative value for the P in T parameter is automatically set (48). Better saying, the same evidence used by children acquiring consistent null-subject languages used to set the D in T parameter (i.e., null subject pronouns) can be used to set the P in T parameter. As already noted, there is no evidence of parameter missetting in children acquiring consistent null-subject language. To capture this, I suggest that there is an immediate setting of D in T and P in T parameters or that these parameters are set very early in the child's development, during a period which could not be detected by studies on the acquisition of null subjects.

(48) D in T (+) —> P in T (-)

Consider, now, the ϕ-dependent parameter. The negative value of the parameter (ϕ-independent) entails that [Spec, TP] can be filled by the goal of T's probing *or* by another element. For example, in Spanish, a ϕ-independent language, [Spec, TP] can be filled by the goal of T's probing, that is, the nominative subject, as seen in (49a), or by the indirect object, which is assigned dative case, as seen in (49b).

(49)a. Juan tiene bombones. *Spanish*

Juan have:3SG chocolates

'Juan has chocolates.'

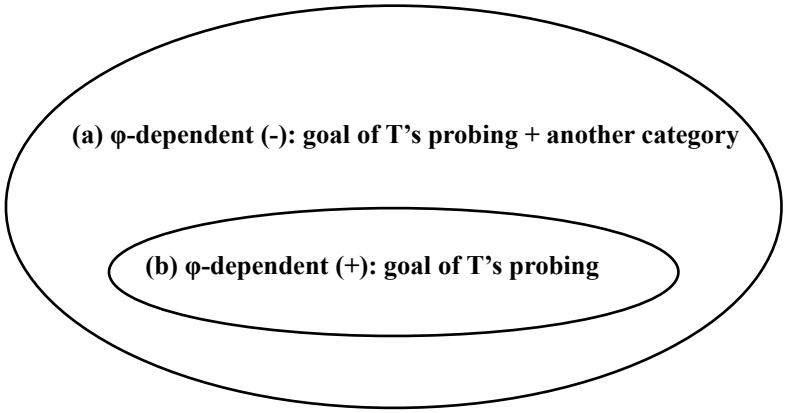
b. A Juan le gustan los chocolates.

to Juan CL like:3PL the chocolates

'Juan likes chocolates.'

The negative value for the ϕ-dependent parameter (ϕ-independent) is in a superset relation with the positive value for the ϕ-dependent parameter, as is represented in Figure 4:

Figure 4: ϕ-dependent parameter and the Subset Principle



Based on Figure 4, the default for the ϕ-dependent parameter is presumably the *positive* value (the subset in Figure 4). The evidence in the input for the child to switch from a positive ((b) ϕ-dependent) to a negative ((a) ϕ-independent) value can vary according to the language. In Spanish, for example, the presence of indirect objects as a subject (as seen in (54b)) indicates that the language is ϕ-independent. In other words, the evidence to switch the ϕ-dependent parameter from positive to negative can be, potentially, any structure that shows the child that [Spec, TP] can be filled by a category other than the goal of T's probing. A child acquiring

Italian or EP would already start out with the appropriate value for the target language (ϕ-dependent (+)).

The table below gives the initial values of the D in T, P in T and ϕ-dependent parameter and the triggers that can potentially be used by the child to set the parameters in *consistent null-subject languages*.

Table 3: Initial values of parameters and potential input evidence for their setting (consistent null-subject languages)

Parameter	***Initial value***	***Input evidence***	***Final value***
D in T	-	Non-restricted definite null subject pronouns.	+
P in T	+	Set with D in T	-
Φ-dependent	+	Structures which show that [Spec, TP] can be filled by a category other than the goal of T's probing (e.g., fronted indirect objects in Spanish).	varies

3.5 Partial null-subject languages: D in T, P in T and φ-dependent Parameters

Considering the initial values in Table 3, deduced from the Subset Principle, the child acquiring the partial null-subject languages discussed here (i.e., BP and Finnish) has to switch the values of both the P in T and the ϕ-dependent parameters, as summarized in the table below:

Table 4: Initial values of parameters and the final values (partial null-subject languages)

Parameter	***Initial value***	***Final value***
D in T	-	-

P in T	+	-
Φ-dependent	+	-

Partial null-subject languages allow definite null subjects in embedded clauses controlled by an argument in a higher clause (Rodrigues 2004, Holmberg 2005, Modesto 2000a, Holmberg and Sheehan 2010)[71] , as seen in the examples below in BP and Finnish.

(50)a. O Pedro$_i$ diz que e_i sabe inglês muito bem. *BP*
The Peter say:3SG that know:3SG English very well
'Peter says that he knows English very well.'

b. Pekka$_i$ väittää [että e_i puhuu englantia hyvin]. *Finnish*
Peter claim:3SG that speak:3SG English well
'Peter claims that he speaks English well.'

(Holmberg 2005: 539)

Null subjects in embedded clauses could serve as a clue to the child acquiring partial null-subject languages that [Spec, TP] does not need to be pronounced. In non-null-subject

[71] As discussed in Chapter 2, section 1.3, there is a debate in the literature whether the null subject in (50) is PRO or not. Rodrigues (2004) Ferreira (2004) and Modesto (2000a) take the position that the null subject in (50) is PRO, but Holmberg and Sheehan (2010) argue that this null subject is not PRO, given that it is not regulated by the control calculus proposed by Landau (2004). If the null subject in (50) is in fact PRO, children would have to misanalyze it as *pro* if null subjects with finite verbs are used to set the P in T parameter. This is so because PRO occurs in every language, even those with a positive value for P in T (e.g., English).

languages (51) and in Icelandic (a semi null-subject language) (52), null subjects in embedded clauses result in an ungrammatical construction. Notice that non-null subject languages and Icelandic are the only languages in Table 2 with a positive value for the P in T parameter.

(51) Peter claims that *(he) speaks Portuguese very well. *English*

(52) Aðeins hann hélt að *(hann) talaði íslensku. *Icelandic*
Only he thought:3SG that. *(he) talked:3SG.SUBJ Icelandic
'Only he thought he could speak Icelandic.'

(Wurmbrand 2017: 344)

In BP, sentences with null quasi-expletives (53) could also be used by the child to set the P in T parameter. The child acquiring Finnish, though, cannot use null expletives or null quasi-expletives as evidence to set the P in T parameter, as in Finnish these elements are overt whenever there is no other category available to check the EPP (Holmberg 2005).

(53) Está chovendo.
Be:3SG raining
'It is raining.'

One might wonder how the child knows that there is a null quasi-expletive in (53). If the child concludes, based on the input, that the verb *estar* typically requires an external argument, she is able to conclude that there is a null subject in (53), despite its lack of semantic content.

Regarding the ϕ-dependent parameter, notice that constructions with a generic null subject could inform children acquiring partial null-subject languages that these languages are ϕ-independent. Among the languages listed in Table 2 which allow generic null subjects, namely, BP, Finnish and Icelandic, all them have a negative value for the ϕ-dependent parameter. That follows from generic null subjects being located in [Spec, *v*P]. Assuming that generic null subjects are located below [Spec, TP], languages with generic null subjects will necessarily allow elements such as adverbials in [Spec, TP] (in the proposal adopted here, [Spec, TP] can sometimes also be empty in sentences with generic null subjects). Therefore, the presence of generic null subjects in a language implies a negative value for the ϕ-dependent parameter:

(54) Generic null subjects —> ϕ-dependent (-)

Notice that this is a one-way implication, though: Spanish is a ϕ-independent language, but it does not have generic null subjects. That is, it is not true that a negative setting for the ϕ-dependent parameter necessarily leads to the presence of generic null subjects in a language.

The table below gives again the initial values of the D in T, P in T and ϕ-dependent parameters and adds the evidence in the input that could potentially be used by the child to set the parameters in *partial null-subject languages*.

Table 5: Initial values of parameters and potential input evidence for their setting (partial null-subject languages)

Parameter	***Initial value***	***Input evidence***	***Final value***
D in T	-	N/A	-
P in T	+	Embedded null subjects Null quasi-expletives (BP)	-
Φ-dependent	+	Generic null subjects.	-

3.6 Semi null-subject languages: D in T, P in T and φ-dependent Parameters

Semi null-subject languages are languages which license only null non-definite pronouns. This includes null expletives and generic null subjects. Icelandic licenses null expletives in weather verbs (quasi-argumental null subjects) and generic null subjects. Kriyol has null expletives in general, although it does not have null generic subjects.

Icelandic is sometimes classified as a partial null-subject language, as in Holmberg (2010a). The reason why Icelandic is not being classified as a partial null-subject language in this paper (following Biberauer 2010), despite the fact that it has generic null subjects, is because the language does not have controlled null subjects in finite clauses, differently than BP and Finnish, which are classified as partial null-subject languages here.

Notice in Tables 6 and 7 that the final values for the parameters here discussed differ in Icelandic and Kriyol:

Table 6: Initial values of parameters and the final values (Icelandic)[72]

Parameter	***Initial value***	***Final value***
D in T	-	-
P in T	+	+
Φ-dependent	+	-

Table 7: Initial values of parameters and the final values (Kriyol)

Parameter	***Initial value***	***Final value***
D in T	-	-
P in T	+	-
Φ-dependent	+	+

As can be seen in the tables above, while the child acquiring Icelandic has to set the value for the ϕ-dependent parameter, the child acquiring Kriyol has to set the value for the P in T parameter. Let's first discuss the setting of the ϕ-dependent parameter by children acquiring Icelandic.

In Icelandic, sometimes [Spec, TP] will be filled by the DP or pronoun which values the T's uϕ-feature and is assigned nominative case, but not always. Sometimes, [Spec, TP] is filled by oblique subjects in Icelandic, if they are the closest category to T, as the sentence in (55) shows:

[72] There is a potential problem with the setting of the P in T parameter if Icelandic allows [Spec, TP] to be empty in constructions with generic null subjects, as BP does. This would change the final setting of the P in T parameter in Icelandic from positive to negative.

(55) Mér voru gefnar peninga. *Icelandic*

Me were given money:PL

'I was given money.'

(Holmberg 2010: 107)

In (55), 'peninga' is the nominative subject, but [Spec, TP] is filled by 'mér', a dative subject. As [Spec, TP] can be filled by an oblique subject in Icelandic, the language is ϕ-independent. Notice that constructions with generic subject also indicate that the language has the negative setting for the ϕ-dependent parameter. In the sentence in (56), an adverbial appears in [Spec, TP] and the generic null subject is in [Spec, *v*P]:

(56) Nú má *e* fara að dansa. *Icelandic*

Now may one go to dance

'One may begin to dance now'

(Sigurðsson and Egerland 2009: 169)

Children acquiring Icelandic can use constructions such as (55) as evidence to set the ϕ-dependent parameter to negative if they have knowledge of Case. As in partial null-subject languages, generic null subjects can also serve as evidence to set the ϕ-dependent parameter in Icelandic.

It is possible, based on the fact that the negative setting for the φ-dependent EPP is in a superset relation with the positive setting for this parameter, that elements such as quirky subjects are acquired later than nominative subjects.

In section 3.3, it was pointed out that as Kriyol is said to have uφ-features even though the language lacks overt marking of person and number, a child acquiring the language will have to assume that Kriyol is a radical pro-drop language, until the grammar of the language tells her otherwise. The lack of thematic null subjects could function as indirect negative evidence leading the child to a restructuring of the system. The D in T parameter will be set as negative if the child uses the lack of thematic null subjects as a clue she is not acquiring a radical pro-drop language.[73] The other parameters are predicted to have the initial value in Table 5 (i.e., P in T (+), φ-dependent (+)), given the Subset Principle.

In order to achieve the parameter values of their target grammar, children acquiring Kriyol have to set the P in T parameter to its negative value. Definite null subjects in embedded position are not available in Kriyol, but null expletives are (Nicolis 2005, 2008). Null expletives could inform the child acquiring Kriyol that the language has a negative value for the P in T parameter.

The tables below show the initial values of the D in T, P in T and φ-dependent parameters and the evidence in the input that could potentially be used by the child to set the parameters in Icelandic and Kriyol (*semi null-subject languages*).

---

[73] I suggest, temporarily, that children acquiring Kriyol use indirect negative evidence to abandon their initial hypothesis that Kriyol is a radical pro-prop language. However, a deeper study of the grammatical properties of Kriyol could reveal that the child is able to use positive evidence to make this change. I leave the question of whether the child is able to use positive evidence or not in Kriyol to abandon her initial hypothesis for future research.

Table 8: Initial values of parameters and potential input evidence for their setting (Icelandic)

Parameter	***Initial value***	***Input evidence***	***Final value***
D in T	-	N/A	-
P in T	+	N/A	+
Φ-dependent	+	Oblique subjects Generic null subjects	-

Table 9: Initial values of parameters and potential input evidence for their setting (Kriyol)

Parameter	***Initial value***	***Input evidence***	***Final value***
D in T	-	N/A	-
P in T	+	Null expletives	-
Φ-dependent	+	N/A	+

3.7 Non-null-subject languages: D in T, P in T and φ-dependent Parameters

Considering the initial values for the three parameters discussed here, deduced from the Subset Principle, children acquiring non-null-subject languages (here, French and English) do not need to change the value of any parameter, as seen in the table below:

Table 10: Initial values of parameters and the final values (non-null-subject languages)

Parameter	***Initial value***	***Final value***
D in T	-	-

P in T	+	+
Φ-dependent	+	+

English and French have a negative value for the *D in T* parameter, since the adult grammar of these languages does not allow null subjects with a definite reading. The value for the *P in T* parameter is positive in English and French, since [Spec, TP] cannot be empty in the language (e.g., expletives are obligatorily overt). The value for the ϕ-dependent parameter is also positive in English and French, since no other category than DPs and pronouns valuing the T's uϕ-features can appear in [Spec, TP]. If the child acquiring non-null-subject languages knows all that, how do we account for the phenomenon of missing subject in these languages (discussed in section 1.1)?

I propose that missing subjects in non-null-subject languages do not result from the missetting of any of the parameters discussed so far. In what follows, I suggest that missing subjects result from subject ellipsis, that is, a phenomenon analogous to subject drop in casual speech or abbreviated registers in non-null-subject adult languages (Bromberg and Wexler 1995; Rizzi 1994; Rizzi 2005). The discussion will focus on missing subjects in English.

3.8 Subject Ellipsis in English

3.8.1 Subject Ellipsis in Adult English

Subject ellipsis is a phenomenon that occurs in finite clauses in English and other non-null-subject languages (e.g., French). Such subject in adult language is called 'diary drop', since the phenomenon is common in abbreviated registers, as in diaries, postcards and text messages.

I am assuming that subject ellipsis is a PF-phenomenon. That is, a full structure is sent to the LF and PF interface. At LF, this structure is interpreted. At PF, the subject of this structure is deleted, since it contains redundant information which can be pragmatically recovered.

According to Nariyama (2004), subject ellipsis is a common phenomenon in casual spoken English. Pragmatically, subject ellipsis occurs when the reference of the subject is recoverable in the context or in a previous sentence. Subject ellipsis in finite clauses in adult English is restricted to coordinate structures and sentence initial position (57), just like subjectless sentences in child English. That is, in non-null-subject languages, subjectless sentences in finite clauses do not occur in the environments listed in (58) in the speech of children or adults (Guasti 2002)[74]:

(57)a. No further moves on the PhD but (I) would like to get into it soon, although the logistics are not easy.

b. (I) should've known better.

(Nariyama 2004: 246-7)

(58)a. after wh-elements

b. in subordinate clauses

c. after a fronted XP other than the subject

[74] As pointed out earlier (see note 51), Roeper and Rochrbacher (2000) and Bromberg and Wexler (1995) found that null subjects occur with wh-questions in non-finitive clauses in child English (i.e., bare verbs in English). As I will be discussed in the next section, null subjects in English can only be analyzed as resulting from subject ellipsis when they occur in finite clauses.

Nariyama (2004) conducted a corpus analysis in order to investigate various syntactic, semantic and pragmatic features of subject ellipsis in English (particularly, Australian English). The author found that the majority of subject ellipses occurs with first person, but they can also occur with other persons mainly when triggered by anaphora and conventional expressions (e.g., "gotta go", "could be", "see you later"). Nariyama analyzed transcriptions of TV dramas, spoken conversation and three casual letters. Importantly, as the author shows, subject ellipsis is not restricted to written abbreviated registers, but is fairly common in spoken languages. Nariyama (2004) did not calculate the percentage of subjectless sentences relative to the total number of utterances. Therefore, the data do not give us information on the frequency of subjectless sentences in adult English, so it cannot be used to establish a comparison with the child data. However, one cannot say that subjectless sentences are rare in spoken English, since the author found subjectless sentences throughout all the corpora analyzed: 53 subjectless utterances were found in 'TV dramas', 49 were found in 'spoken conversation' and 20 in 'casual letters'.

According to Nariyama (2004), pragmatically, subject ellipsis is triggered when the speaker believes, consciously or not, that the addressee can infer the intended identity of the null subject. This specifically happens in the following linguistic contexts: (i) when the reference of the elided subject can be recovered by an anaphoric relation relying on linguistic context, (ii) by deixis, (iii) when the elided subject is a dummy subject and (iv) in conventional expressions.

3.8.2 Subject ellipsis in Child English

The parallelism between early null subjects in finite clauses and adult subject ellipsis in non-null-subject languages consists in the similar syntactic context in which these phenomena

occur: such null subjects generally only occur sentence initially. This syntactic parallel has been the instigator of proposals arguing that early null subjects are an instance of diary drop (Bromberg and Wexler 1995; Rizzi 1994).

As early null subjects in children are manifested in spoken language, one can try to connect subject-drop in casual speech and early subject-drop. Haegeman and Ihsane (2001) argue against the analysis of missing subjects as analogous to subject-drop in casual speech. The authors observe that while articles in spoken abbreviated register can be omitted only in initial position, children omit articles clause internally. Similarly, while auxiliary and copula deletion are confined to the initial position of the sentence in spoken abbreviated register, in child language omission of auxiliary and copula occur more generally. One could argue, though, that article, auxiliary and copula deletion in child language are not related to missing subjects and thus not analogous to article, auxiliary and copula deletion in adult language. Indeed, these phenomena in child language can be analyzed as unrelated to subject ellipsis.

Regarding the omission of articles, Guasti et al. (2008) analyze the production of articles by children speakers of Catalan, Italian and Dutch. The authors found that for children speaking Italian and Dutch there is an initial stage in which no articles are used (it is not clear whether this phase is absent in Catalan or not). This is explained by the hypothesis that children start out by assuming that their language has no article at all. In the next stage, children speaking Catalan, Italian and Dutch might omit articles, although Romance learners soon discover that in their language NPs cannot occur in a sentence without a DP: this is triggered by the discovery that the article paradigm is complete in Romance languages. Dutch-speaking children take longer to stop omitting articles because the presence or absence of articles varies according to the lexical item.

During the second stage, Dutch-speaking children seem to misclassify some countable nouns as mass nouns (e.g., they classify 'apple', a countable noun, as a mass noun such as 'furniture'), which leads to the omission of articles. Importantly, in this proposal, article deletion is not restricted to the beginning to a sentence: article omission can occur anywhere in the sentence. That is, this proposal, which does not relate article omission to subject omission, can account for the pattern of article omission seen in child language.

Auxiliary and copular deletion can also be analyzed as a phenomenon not associated with subject ellipsis, but with the Optional Infinitive phase, which makes children produce null subjects in non-finite clauses (i.e., bare verbs in English). Missing subjects in non-finite clauses occur in different environments than missing subjects in finite clauses. English-speaking children produce null subjects in wh-questions with bare verbs, but not with finite verbs. Roeper and Rohrbacher (2000) found that 95% of the null subjects in wh-questions occur with bare verbs and only 5% of these null subjects occur with finite verbs. Bromberg and Wexler (1995) report similar results. As null subjects in finite and non-finite clauses do not occur in the same environment in the speech of children acquiring English, it can be argued that while null subjects in finite clauses result from subject ellipsis, null subject in non-finite clauses result from another phenomenon. As Valian (2016: 401) observes null subjects in non-finite clauses (RIs) exhibit a different developmental course than null subjects in finite clauses, with RIs persisting for longer than null subjects in finite clauses. That is, subject ellipsis can be one of the sources of early missing subjects, while null subjects in non-finite clauses result from a different source.

I follow Bromberg and Wexler (1995) in assuming that null subjects co-occuring with bare verbs are an instance of PRO, thus they are not restricted to the initial position of the

sentence, differently than missing subjects resulting from subject ellipsis. The omission of auxiliary and copula is related to the Optional Infinitive stage, a stage in which TENSE is omitted (Wexler 1992).

The frequency of overt subjects increases over time in the speech of children acquiring English. Boster (1997) analyzed the spontaneous speech of three English-speaking children: Adam, Eve and Peter. In the first ages analyzed the proportion of subject omission in finite clauses in the speech of these three children ranges from 100% to 70% and it reaches 5% to 0% in the last ages analyzed (Boster 1997: 132).

Since English-speaking children have the same syntax as adults, which allows null subjects in finite clauses only sentence initially, how could we explain the decreasing use of null subjects over time? One possibility is that children need to learn the pragmatic context in which subject ellipsis is possible in English. As discussed, English-speaking adults allow subject ellipsis when it is clear for the speaker that the addressee can infer the intended identity of the null subject. This specifically happens when (i) the reference of the null subject can be recovered by an anaphoric relation relying on linguistic context, (ii) by deixis, (iii) when the null subject is a dummy subject and (iv) in conventional expressions (e.g., “gotta go”, “see you”). I leave for future work the investigation of whether children allow null subjects in pragmatic contexts not allowed by adults (i.e., whether children make commission errors). Another possibility is that as the child’s language is tied to the “here and now” there are many uses of null subjects recovered by deixis, a pragmatic context in which adults allow subject ellipsis, but that occurs more rarely in adult conversation, since the speech of adults is less tied to the “here and now” (i.e., adults often talk about the past, future etc.) than the child’s speech. Notice that this second analysis

would be convergent with the results presented by Simões (1999) for BP, who found that the somewhat high frequency of null subjects used by the child she analyzed can be explained by the high occurrence of null subjects used in "here and now" contexts (i.e., null subjects recovered by the physical presence of the referent). These two hypotheses are left to be investigated in the future.

It remains to be answered why children acquiring consistent null-subject languages, as Italian, omit subjects roughly in the same proportion as adults (Lorusso et al. 2004 and Serratrice 2005) if the high proportion of subject ellipsis in child English compared to adult English can be explained due to pragmatic deficits or by the fact that the child's speech is tied to the "here and now". That is, why would not the same pragmatic deficit or tendency to tie the speech to the "here and now" cause Italian-speaking children to produce more null subjects than adults? Under the account outlined here, children acquiring any (non-)null-subject languages obey the same syntactic restrictions as adults, but pragmatics or the context in which speech is produced varies between adults and children. In consistent null-subject languages, null subjects occur in most contexts and more frequently than overt subjects, which makes pragmatic deficits or differences in the speech context hard to detect when the production of null subjects is analyzed. That is to say, children acquiring consistent null-subject languages presumably also have pragmatic deficits or have their speech tied to the immediate context, but as most of the utterances should have null subjects in their languages, it is likely that null subjects produced by children acquiring these languages are normally expected in the adult grammar.[75]

[75] Another possibility is that subject ellipsis is blocked in consistent null subject languages (see Bošković 2018).

4 Conclusion

In this chapter, I proposed an analysis of how children set the null subject parameters in consistent null-subject languages, partial null-subject languages, non-null subject languages and two semi null-subject languages (Icelandic and Kriyol). The proposal is based on Holmberg's (2010a) D in T, P in T and ϕ-dependent parameters.

The initial setting of the parameters was deduced from the Subset Principle (Berwick 1982; Wexler and Manzini 1987). According to this principle, initially children should choose the most restrictive value for a parameter, which corresponds to the subset. If the child's initial choice does not account for the data provided by the input, children can switch to the most unrestrictive value for the parameter (the superset) based on positive evidence. The initial value of the parameters in the analysis conducted in section 3 is the following:

Table 11: Initial values of parameters

Parameter	***Initial value***
D in T	-
P in T	+
Φ-dependent	+

I proposed that the evidence in the input which can be used by children acquiring consistent null-subject languages to switch the value of the D in T parameter from negative to positive is the presence of unrestricted definite null subjects in the language. As was discussed, although definite null subjects are also possible in BP, a partial null-subject language, they only occur when the reference of the null subject can be recovered by the topic or by the physical

presence of the referent. Once the child acquiring a consistent null-subject language switches the value of the D in T parameter from negative to positive, they will also switch the initial value of the P in T parameter, as these two parameters are in a bleeding relation. Children acquiring Spanish, additionally, need to switch the value of the ϕ-dependent parameter from positive to negative. It was proposed that fronted indirect objects could function as evident to set the ϕ-dependent parameter in Spanish.

Children acquiring partial null-subject languages, such as BP and Finnish, need to switch the values of the P in T and ϕ-dependent parameters to negative. Embedded null subjects were suggested as input evidence to reset the P in T parameter both in BP and in Finnish. In BP, null quasi-expletives could also serve as an indicator that P in T is negative in the language. It was proposed that the evidence to switch the value of the ϕ-dependent parameter from positive to negative is the presence of generic null subjects in the language.

Importantly, as there is no evidence of parameter missetting in the grammar of children acquiring consistent null-subject languages and partial null-subject languages, the setting of the parameters must occur very early in the child's development.

The two semi null-subject languages here analyzed differ from each other regarding the final setting of the parameters. Although both languages are classified as semi null-subject languages, as they lack definite null subjects, they have distinct grammars to be acquired. A child acquiring Icelandic needs to switch the value of the ϕ-dependent parameter, while the child acquiring Kriyol needs to switch the value of the P in T parameter. In Icelandic, potential clues to set the ϕ-dependent parameter are generic null subjects and constructions with oblique subjects. In Kriyol, expletive null subjects could function as a positive evidence for the child to switch the value of the P in T parameter to negative.

Children acquiring non-null-subject languages do not need to reset the initial value of these parameters, as they correspond to the final values in the target languages. We were left with the question of why children omit subjects in non-null-subject languages if the initial value of the D in T, P in T and ϕ-dependent parameters correspond to their target language. Focusing on missing subjects in finite clauses in English, I proposed that they result from subject ellipsis, an option available in the grammar of adults. Under this account, the initial high proportion of subjectless sentences in the child's speech can be explained by two hypotheses: (i) the child needs to learn in which pragmatic context subject ellipsis is allowed in their language or (ii) the child produces a high number of null subjects which are recovered by the presence of the referent in the immediate context, a pragmatic situation where subject ellipsis is possible in adult language, albeit not common, as the adult's speech is not tied to the "here and now".

Chapter 7 - Conclusions and Future Research

In this thesis, I discussed the syntactic structure of null impersonals with generic reading in Brazilian Portuguese (BP) and investigated its acquisition. I also proposed a learning model to account for the pattern of null subjects in BP and in different (non-)null subject languages (i.e., non-null-subject languages, consistent null-subject languages, partial null-subject languages, semi null-subject languages and radical pro-drop languages).

In this concluding chapter, I return to the research questions raised in Chapter 1 and summarize the answers I proposed to these questions.

I. What is the syntactic structure of null impersonals with generic reading in BP? Which constraints are imposed on the well-formedness of these sentences?

Following Roberts (2010) and Holmberg (2010a), I assumed that null impersonals with generic reading have a φP pronoun (that is, a pronoun with the features [3SG, uCase]) which enters into an Agreement relation with T and is deleted by chain reduction. The generic reading of this pronoun is explained by the absence of a D-feature in T. Consistent null-subject languages (e.g., Italian, European Portuguese), which have 3rd person definite null subjects in matrix clauses, also have a φP pronoun, but they have a D-feature in T, which is responsible for the definite reading of φP pronouns, as represented in (1). As BP and other partial null-subject languages (e.g., Finnish) do not have a D-feature in T, the φP pronoun can only have the generic reading, as represented in (2).

(1) a. [T, D, uφ, NOM] [vP [3SG, uCase] v…] → *Consistent null-subject languages*

b. [T, D, 3SG, NOM] [*v*P [3SG, NOM] *v*...] →

c. [T, D, 3SG, NOM] [*v*P [~~3SG, NOM~~] *v*...]

(2) a. [T, uφ, NOM] [vP [3SG, uCase] v…] → *Partial null-subject languages*

b. [T, 3SG, NOM] [*v*P [3SG, NOM] *v*...] →

c. [T, 3SG, NOM] [*v*P [~~3SG, NOM~~] *v*...]

I proposed that an overt marker of genericity is required for complete well-formedness of generic impersonal structures in BP. Overt markers of genericity can assume a variety of forms, but I focused on overt deontic modals and on the *se* impersonal clitic pronoun. Contrary to Holmberg (2005, 2010a,b), I claimed that in null impersonal structures there is no EPP-feature which needs to be checked by elements such as fronted adverbials. This seems to be true at least for BP.

II. Based on empirical evidence, can we conclude that impersonal structures in BP (i) have an EPP feature which needs to be checked by an element other than the null generic pronoun? Can we also conclude, based on empirical evidence, that (ii) the generic reading of impersonal sentences requires an overt marker of genericity?

In order to answer this research question, I conducted an online survey with native speakers of BP. Regarding (i), the results suggest that there is no EPP feature which needs to be checked in impersonal structures, given the acceptability of sentences like (3), where there is no

element that could satisfy the EPP. Generic ϕP pronouns are unable to satisfy the EPP, as they are non-head members of an argument chain headed by T, located in [Spec, vP].

(3) Não pode largar a escola.
Not can drop:INF the school
'One should not drop out of school.'

Regarding (ii), the results suggested that BP-speakers prefer generic impersonal structures with an overt marker of genericity, but not that they require it. This was shown by the fact that in sentences without a deontic modal, the presence of the generic clitic *se* improved the acceptability of these sentences.

III. Do children acquiring BP know that null subjects in impersonal constructions have the generic reading rather than the definite one? That is to say, do BP-speaking children know they are acquiring a partial null-subject languages instead of a consistent null-subject language?

In order to answer this question, I conducted an experiment using the Truth-Value Judgement Task as a method. If BP-speaking children understand that the null subject in sentences like (4) have the generic reading instead of the definite one, we predicted that they should accept the generic reading of the sentence and reject its definite reading.

(4) Nessa escola não *e* pode escovar os dentes depois de comer.
In.this school not can brush:INF the teeth after of eat:INF

'In this school one cannot brush one's teeth after eating.'

The results showed that children as young as 4-years-old exhibit sensitivity to the adult grammar, correctly rejecting the definite reading and accepting the generic reading of the null subject in impersonal structures. That is, at a very young age, BP-speaking children know that BP is a partial null-subject language and not a consistent null-subject language such as EP.

IV. In which contexts do children allow null impersonals in BP? Do they obey the same restrictions imposed by the adult grammar?

In order to investigate if children obey the same restrictions as adults for the well-formedness of impersonal sentences, I conducted a Felicity Judgement Task. Children had to judge which puppet said the test sentence in the "best way". Two puppets alternated between uttering sentences with *se* and with a null generic pronoun.

Of particular interest were the results for the conditions ADV (5) and (ADV+D) (6). The expected answer for (5) was *se*, as BP-speakers prefer impersonal structures with an overt generic marker. The expected answer for (6) was 'both' (both *se* and a null subject are acceptable), since, as concluded in Chapter 3, nothing prevents the co-occurrence of two overt markers of genericity in impersonal structures in BP (a deontic modal and the clitic *se*).

(5) Nesse escola (se) traz brinquedo. *ADV*

In.this school (SE) bring:INF toy

'In this school one brings toys.'

(6) Nessa escola não (se) pode comer banana. *ADV+D*

In.this school not (SE) can eat:INF banana

'In this school one cannot eat bananas.'

7-year-olds and 6-year-olds exhibited sensitivity to the adult grammar, correctly identifying the restrictions imposed on the structures above. 5-year-olds and 4-year-olds exhibited worse performance than the other groups. There was a tendency among children of these younger age groups to always choose sentences with a null subject. It seems that at the age of 5-years-old children did not completely acquire the impersonal pronoun *se* which would explain the pattern of answers of the youngest age groups. *Se* is not a form widely present in the input and it is possible that this form develops with schooling. That is not to say that *se* is not part of the early grammar, but instead that school can promote the development of grammatical aspects which are part of the internal grammar, but not frequent in the input.

V. How do children acquire different (non-)null-subject grammars?

In order to address this learnability problem, I proposed a learning model based on Holmberg's (2010a) D in T, P in T and ϕ-dependent parameters.

I deduced the initial setting of these parameters from the Subset Principle (Berwick 1982; Wexler and Manzini 1987), reaching the following initial value for the Null Subject Parameters:

Table 1: Initial values of parameters

Parameter	***Initial value***
D in T	-

P in T	+
Φ-dependent	+

Following that, I proposed which evidence the child acquiring different (non-)null-subject languages could use to set the values of the Null Subject Parameters when they do not correspond to their initial values in the child's language.

Some issues were left for future research. Below I list the limitations of this thesis and the questions that should be investigated in the future:

I. Why do BP-speakers prefer generic impersonal structures with overt markers of genericity? Why is a law-like background not enough for the complete well-formedness of generic impersonal sentences?
II. Conduct an empirical study to test whether null impersonals can have the existential reading in BP. Particularly, investigate whether habitual sentences such as (7) can have the existential reading (as claimed by Carvalho 2018, 2019):

(7) Nessa loja *e* vende saia.
In.this store sell:3SG skirt
'In this store someone sells skirts.'

III. If null impersonals can have the existential reading for adults, how do children acquire impersonals with the existential reading?

IV. Can the modal *tem que* (have to) co-occur with *se*? If not, as suggested by the survey discussed in Chapter 3, which properties are responsible for this restriction?

V. Experiment 2, which investigated whether children obey the same restrictions as adults in generic impersonal sentences had only one test sentence per condition. Therefore, it is necessary to conduct a study with more test sentences per condition and see if the results are replicated.

VI. Impersonals with wh-questions seem to license a generic operator, but it is unclear how the wh-element would function as an overt marker of genericity. This question needs to be investigated in future work.

VII. In the learning model I proposed in Chapter 6, the evidence to set the Null Subject Parameters comes from the structure itself (e.g., the D in T parameter is set based on whether a language has unrestricted definite subjects in matrix clauses or not). A study looking for additional evidence should be conducted in the future.

VIII. In the same learning model, children acquiring Kriyol are proposed to use indirect negative evidence (i.e., absence of thematic null subjects) to tell them that Kriyol is not a radical pro-drop language. A deeper investigation on the grammar of this language should be conducted to see whether children acquiring Kriyol can use positive evidence to reach the conclusion that Kriyol is not a radical pro-drop language.

IX. Conduct a corpus study to investigate whether English-speaking children use null subjects in finite clauses in pragmatic contexts not allowed by adults, as predicted by the hypothesis that the large number of null subjects in finite clauses is at least partially explained by lack of pragmatic knowledge children have of when subjects can be dropped.

X. Conduct a corpus study to investigate whether the high occurrence of null subjects in finite clauses by English-speaking children can be at least partially explained by the large number of utterances produced in "here and now" contexts.

Although definitive analysis must await the completion of the additional research suggested above, on the basis of the research in this dissertation it can be concluded that (i) in BP complete well-formedness of null impersonals with generic reading depends on an overt marker of genericity (such as *se* and a deontic modal); (ii) there is no EPP feature which needs to be checked in null impersonals with generic reading in BP; (iii) children as young as 4-years-old know that BP is a partial null-subject language, correctly rejecting the definite interpretation of null subjects in impersonal structures and accepting its generic reading; (iv) children start out with the most restrictive grammar regarding the use of null subjects: a non-null-subject grammar.

References

Alexiadou, A., & Anagnostopoulou, E. (1998). Parametrizing Agr: Word Order, V-Movement and EPP-Checking. *Natural Language & Linguistic Theory*, *16*(3), 491–539.

Assis, J. (2017). *De se a (vo)cê: O percurso da indeterminação do Português Brasileiro* [Ph.D. dissertation]. Universidade Federal de Uberlândia.

Barbosa, P. (2010). *Partial pro-drop and null NP-anaphors*. NELS 41, UPenn.

Barbosa, P., Duarte, M. E. L., & Kato, M. A. (2005). Null Subjects in European and Brazilian Portuguese. *Journal of Portuguese Linguistics*, *4*(2), 11.

Berger-Morales, J., & Salustri, M. (2003). Supporting the Separate Systems Hypothesis: A Case Study in Bilingual Acquisition of Italian and German. In A. T. Pérez-Leroux & Y. Roberge (Eds.), *Current Issues in Linguistic Theory* (Vol. 244, pp. 263–280). John Benjamins Publishing Company.

Bertolino, K. (2017). *Modals and Impersonal Null Subjects in Brazilian Portuguese* [Master's thesis]. University of Connecticut.

Berwick, R. (1982). *Locality principles and the acquisition of syntactic knowledge* [Ph.D. dissertation]. Massachusetts Institute of Technology.

Blevins, J. P. (2003). Passives and impersonals. *Journal of Linguistics*, *39*(3), 473–520.

Bloom, L. (1970). *Language development: Form and function in emerging grammars*.

Bloom, L., Lightbown, P., & Hood, L. (1975). *Structure and Variation in Child Language*.

Bloom, P. (1990). Subjectless Sentences in Child Language. *Linguistic Inquiry*, *21*(4), 491–504.

Boeckx, C. (2003). *Intricacies of Icelandic Agreement*, unpublished ms., Harvard University and University of Maryland.

Borer, H. (1986). I-subjects. *Linguistic Inquiry*, *21*, 375–416.

Bošković, Ž. (2002). A-movement and the EPP. *Syntax*, *5*(3), 167–218.

Bošković, Ž. (2007). On the Locality and Motivation of Move and Agree: An Even More Minimal Theory. *Linguistic Inquiry*, *38*(4), 589–644.

Bošković, Ž. (2018). On Pronouns, Clitic Doubling, and Argument Ellipsis: Argument Ellipsis as Predicate Ellipsis. *English Linguistics 35*: 1-37.

Boster, C. (1997). *Processing and parameter setting in language acquisition: A computational approach* [Ph.D. dissertation]. University of Connecticut.

Braine, M. D. S. (1963). The Ontogeny of English Phrase Structure: The First Phase. *Language*, *39*(1), 1.

Brody, M. (2013). Silent people: The pseudo-impersonal. In J. Brandtler, V. Molnár, & C. Platzack (Eds.), *Approaches to Hungarian* (Vol. 13, pp. 33–44). John Benjamins Publishing Company.

Bromberg, H., & Wexler, K. (1995). Null subjects in Wh questions. *MIT Working Papers in Linguistics*, *26*, 221–248.

Camacho, J. (2013). *Null subjects*. Cambridge University Press.

Carvalho, J. (2018). Diferentes tipos de sujeitos nulos no português brasileiro. *ReVEL*, *16*(30), 78–106.

Carvalho, J. (2019). Incorporated subjects in existential impersonal sentences in Brazilian Portuguese. In P. Herbeck, B. Pöll, & A. C. Wolfsgruber (Eds.), *Semantic and syntactic aspects of impersonality*. Buske.

Castilho, A. T. de, & Lopes, C. (Eds.). (2001). O percurso de a gente em tempo real de longa duração. In *História do português brasileiro* (Vol. 2, pp. 127–148). Editora Contexto.
Chierchia, G. (1995). Individual Level Predicates as Inherent Generics. In G. N. Carlson & F. J. Pelletier (Eds.), *The generic book*. University of Chicago Press.
Chomsky, N. (1981). *Lectures on government and binding: The Pisa lectures*. Mouton de Gruyter.
Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. MIT Press.
Chomsky, N. (1995). *The Minimalist program*. The MIT Press.
Chomsky, N. (2001). Derivation by Phase. In K. L. Hale & M. J. Kenstowicz (Eds.), *Ken Hale: A life in language*. MIT Press.
Cinque, G. (1988). On Si-Constructions and the Theory of Arb. *Linguistic Inquiry*, *19*(4), 521–581.
Clahsen, H. (1989). Constraints on Parameter Setting: A Grammatical Analysis of Some Acquisition Stages in German Child Language. *Language Acquisition*, *1*(4), 361–391. JSTOR.
Comparini, A. (2008). A natureza multissignificativa dos verbos modais: Uma análise da língua falada no Brasil. *Revista Científica de Letras*, *4*(4), 31–46.
Costa, A., Costa, A., & Gonçalves, A. (2017). Consciência linguística: Aspectos sintáticos. In M. J. Freitas & A. L. Santos (Eds.), *A aquisição de língua materna e não materna: Questões gerais e dados do Português*. Language Science Press.
Costa, A. L., Alexandre, N., Santos, A. L., & Soares, N. (2008). Efeitos de modelização do input: O caso da aquisição de conectores. *Textos Selecionados Do XXIII Encontro Da Associação Portuguesa de Linguística*, 131–142.
Costa, J. (2004). *Subject positions and interfaces: The case of European Portuguese*. Mouton de Gruyter.
Costa, J., Lobo, M., & Silva, C. (2011). Subject–object asymmetries in the acquisition of Portuguese relative clauses: Adults vs. children. *Lingua*, *121*(6), 1083–1100.
Crain, S., & McKee, C. (1985). The acquisition of structural restrictions on anaphora. In C. Berman & J. McDonough (Eds.), *Proceedings of NELS 15*.
Dahl, Ö. (1975). On generics. In E. L. Keenan (Ed.), *Formal Semantics of Natural Language* (1st ed., pp. 99–111). Cambridge University Press.
Delfitto, D. (2006). Bare Plurals. In M. Everaert & H. van Riemsdijk (Eds.), *The Blackwell Companion to Syntax* (pp. 214–259). Blackwell Publishing.
Diesing, M. (1992). *Indefinites*. MIT Press.
Duarte, I., & Silva, M. C. F. (2016). The Null Subject Parameter and the Structure of the Sentence in European and Brazilian Portuguese. In W. L. Wetzels, J. Costa, & S. Menuzzi (Eds.), *The Handbook of Portuguese Linguistics* (1st ed., pp. 234–253). Wiley.
Duarte, M. E. L. (1993). Do pronome nulo ao pronome pleno: A trajetória do sujeito no português do Brasil. In I. G. Roberts & M. A. Kato (Eds.), *Português brasileiro: Uma viagem diacrônica: Homenagem a Fernando Tarallo*.
Duarte, M. E. L. (1995). *A perda do princípio "Evite Pronome" no Português Brasileiro* [Ph.D. dissertation]. Universidade Estadual de Campinas.
Egerland, V. (2003). Impersonal Pronouns in Scandinavian and Romance. In *Working Papers in Scandinavian Syntax* (Vol. 71). Department of Scandinavian Languages, Lund University.

Faria, I. (1993). A Aquisição da Noção "Agente" e a Produção de Sujeitos Sintáticos por Crianças Portuguesas até os Dois Anos e Meio. *Revista Internacional de Língua Portuguesa*, *10*, 16–50.
Fenger, P. (2018). How impersonal does one get?: A study of man-pronouns in Germanic. *The Journal of Comparative Germanic Linguistics*, *21*(3), 291–325.
Ferreira, Marcelo B. 2004. Hyperraising and Null Subjects in Brazilian Portuguese. In *Collected Papers on Romance Syntax*, ed. by Ana Castro, Valentine Hacquard, Andres Salanova and Marcelo Ferreira, 57-85. MIW Working Papers in Linguistics, Department of Linguistics and Philosophy, MIT, Cambridge, Mass.
Foppolo, F., Guasti, M. T., & Chierchia, G. (2012). Scalar Implicatures in Child Language: Give Children a Chance. *Language Learning and Development*, *8*(4), 365–394.
Frascarelli, M. (2007). Subjects, topics and the interpretation of referential pro: An interface approach to the linking of (null) pronouns. *Natural Language & Linguistic Theory*, *25*(4), 691–734.
Galves, C. (1986). A interpretação "reflexiva" do pronome no português do Brasil. *D.E.L.T.A*, *2*(2), 249–264.
Gerken, L. (1991). The metrical basis for children's subjectless sentences. *Journal of Memory and Language*, *30*(4), 431–451.
Guasti, M. T. (1996). Acquisition of Italian Interrogatives. In H. Clahsen (Ed.), *Language Acquisition and Language Disorders* (Vol. 14, p. 241). John Benjamins Publishing Company.
Guasti, M. T. (2002). *Language acquisition: The growth of grammar* (1. paperback ed). MIT Press.
Guasti, M. T., Gavarró, A., de Lange, J., & Caprin, C. (2008). Article Omission Across Child Languages. *Language Acquisition*, *15*(2), 89–119.
Haegeman, L., & Ihsane, T. (2001). Adult Null Subjects in the non-pro-drop Languages: Two Diary Dialects. *Language Acquisition*, *9*(4), 329–346.
Hale, K. L. (1983). Papago (k)c. *International Journal of American Linguistics*, *49*, 299–327.
Hale, K. L., Kenstowicz, M. J., & Chomsky, N. (Eds.). (2001). Derivation by Phase. In *Ken Hale: A life in language*. MIT Press.
Hamann, C., & Plunkett, K. (1998). Subjectless sentences in child Danish. *Cognition*, *69*(1), 35–72.
Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. [Ph.D. dissertation]. University of Massachusetts.
Hofmann, T. R. (1976). Past Tense Replacement and the Modal System. In J. D. McCawley (Ed.), *Notes from the Linguistic Underground* (pp. 85–100). BRILL.
Holmberg, A. (2003a). Topic-drop or VP-focus. In L.-O. Delsing & C. Platzack (Eds.), *Grammatik i fokus: Festskrift till Christer Platzack den 18 november 2003* (pp. 159–166). Inst. för nordiska språk, Univ.
Holmberg, A. (2003b). Yes/no-questions and the relation between tense and polarity in English and Finnish. In P. Pica (Ed.), *Linguistic Variation Yearbook* (Vol. 3, pp. 43–68).
Holmberg, A. (2005). Is There a Little Pro? Evidence from Finnish. *Linguistic Inquiry*, *36*(4), 533–564.
Holmberg, A. (2010a). Null subject parameters. In T. Biberauer, A. Holmberg, I. Roberts, & M. Sheehan (Eds.), *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge University Press.

Holmberg, A. (2010b). The null generic subject pronoun in Finnish: A case of incorporation in T. In T. Biberauer, A. Holmberg, I. Roberts, & M. Sheehan (Eds.), *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge University Press.
Holmberg, A. (2017). Linguistic Typology. In I. Roberts (Ed.), *The Oxford Handbook of Universal Grammar* (pp. 354–376). Oxford University Press.
Holmberg, A., Nayudu, A., & Sheehan, M. (2009). Three partial null-subject languages: A comparison. *Studia Linguistica*. *56*(1)
Holmberg, A., & Phimsawat, O.-U. (2015). Generic pronouns and phi-features: Evidence from Thai. *Newcastle and Northumbria Working Papers in Linguistics (Second Asian and European Linguistics Conference Proceedings*, *21*.
Holmberg, A., & Sheehan, M. (2010). Control into finite clauses in partial null-subject languages. In T. Biberauer, A. Holmberg, I. Roberts, & M. Sheehan (Eds.), *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge University Press.
Hornstein, N. (1999). Movement and Control. *Linguistic Inquiry*, *30*(1), 69–96.
Huang, C.-T. J. (1984). On the distribution and reference and empty pronouns. *Linguistic Inquiry*, *15*(4), 531–574.
Hyams, N. (1991). A reanalysis of null subjects in child language. In *Theoretical issues in language acquisition: Continuity and change in development* (pp. 249–267). Lawrence Erlbaum Associates.
Hyams, N. (2011). Missing Subjects in Early Child Language. In J. de Villiers & T. Roeper (Eds.), *Handbook of Generative Approaches to Language Acquisition* (Vol. 41, pp. 13–52). Springer Netherlands.
Hyams, N. M. (1986). *Language acquisition and the theory of parameters*. D. Reidel Pub. Co.; Sold and distributed in the U.S.A. and Canada by Kluwer Academic Publishers.
Hyams, N., & Wexler, K. (1993). On the grammatical basis of null subjects in child language. *Linguistic Inquiry*, *24*, 421–459.
Jackson-Maldonado, D., Maldonado, R., & Thal, D. J. (1998). Reflexive and middle markers in early child language acquisition: Evidence from Mexican Spanish. *First Language*, *18*(54), 403–429.
Jaeggli, O., & Safir, K. J. (1989). The Null Subject Parameter and Parametric Theory. In O. A. Jaeggli & K. J. Safir (Eds.), *The Null Subject Parameter* (Vol. 15, pp. 1–44). Springer Netherlands.
Jaggar, P. J. (2001). *Hausa*. John Benjamins Publishing Company.
Jakobson, R. (1990). The sound laws of child language and their place in general phonology. In L. R. Waugh & M. Monville-Burston (Eds.), *On language*. Harvard University Press.
Jelinek, E. (1984). Empty categories, case, and configurationality. *Natural Language and Linguistic Theory*, *2*(1).
Jovović, I. Serbo-Croatian pronouns are topic-shift anaphors. FASL 29 (*online conference*).
Júnior, F. (1882). *Como se fazia um deputado*. Teatro Recreio Dramático.
Kang, J. (2014). On the Absence of TP and its Consequences: Evidence from Korean. [Ph.D. dissertation]
Kato, M. A. (1999). Strong and weak pronominals in the null subject parameter. *Probus*, *11*(1), 1–38.
Kiss, K. É. (Ed.). (1995). *Discourse configurational languages*. Oxford University Press.
Kratzer, A. (1995). Individual-Level and Stage-Level Predicates. In G. N. Carlson & F. J. Pelletier (Eds.), *The generic book*. University of Chicago Press.

Krifka, M., Pelletier, F. J., Carlson, G. N., ter Meuler, A., Godehard, L., & Chierchia, G. (1995). Genericity: An introduction. In G. N. Carlson & F. J. Pelletier (Eds.), *The generic book*. University of Chicago Press.
Kunnari, S., Torn-Leesik, R., & Tolonen, A.-K. (2010). *The acquisition of "short passive" by Finnish and Estonian 4- and 5-year-old children* [Poster]. COST A33 Final Conference "Let the children speak: learning of critical language skills around 25 languages," London.
Laitinen, L. (2006). Zero person in Finnish: A grammatical resource for construing human reference. In M.-L. Helasvuo & L. Campbell (Eds.), *Current Issues in Linguistic Theory* (Vol. 277, pp. 209–231). John Benjamins Publishing Company.
Landau, I. (2004). The Scale of Finiteness and the Calculus of Control. *Natural Language and Linguistic Theory*, *22*(4), 811–877.
Lemle, M. (1985). Pronomes, anáforas, zero: Observação sobre uma mudança linguística. *D.E.L.T.A*, *1*, 121–124.
Lopes, R. (1999). *Uma proposta minimalista para o processo de aquisição da linguagem*. Universidade Estadual de Campinas.
Lopes, R. (2003). The production of subject and object in Brazilian Portuguese by a young child. *Probus*, *15*(1).
Lorusso, P., Caprin, C., & Guasti, M. T. (2005). *Overt Subject Distribution in Early Italian Children*. BUCLD 29.
Lucchesi, D., Baxter, A. N., & Silva, J. (2009). A concordância verbal. In I. Ribeiro, A. N. Baxter, & D. Lucchesi (Eds.), *O português afro-brasileiro*. (pp. 331–371).
Lunguinho, M. (2014). On the acquisition of root and epistemic modals in Brazilian Portuguese. *ReVEL*, *Special Issue 8*, 131–159.
Magalhães, T. (2006). *O sistema pronominal sujeito e objeto na aquisição do portugues europeu e do portugues brasileiro* [Ph.D. dissertation]. Universidade Estadual de Campinas.
Manninen, S., & Nelson, D. (2004). What is a passive? The case of Finnish. *Studia Linguistica*, *58*(3), 212–251.
Maratsos, M. (1982). The child's construction of grammatical categories. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art*. Cambridge University Press.
Mendikoetxea, A. (2008). Clitic impersonal constructions in Romance: Syntactic features and semantic interpretation. *Transactions of the Philological Society*, *106*(2), 290–336.
Modesto, M. (2000a). *On the identification of null arguments* [PhD dissertation]. University of Southern California.
Modesto, M. (2000b). Null subjects without "rich" agreement. In M. A. Kato & E. V. Negrão (Eds.), *Brazilian Portuguese and the Null Subject Parameter* (pp. 147–174). Iberoamericana Vervuert.
Moltmann, F. (2007). Generic one, arbitrary PRO, and the first person. *Natural Language Semantics*, *14*(3), 257–281.
Nariyama, S. (2004). Subject ellipsis in English. *Journal of Pragmatics*, *36*(2), 237–264.
Nicolis, M. (2005). *On pro-drop* [Ph.D. dissertation]. Università degli Studi di Siena.
Nicolis, M. (2008). The null subject parameter and correlating properties: The case of Creole languages. In T. Biberauer (Ed.), *The limits of syntactic variation* (Vol. 132, pp. 271–294). John Benjamins Publishing Company.
Nunes, C. (2015). Estudo diacrônico da colocação pronominal em complexos verbais na escrita do PB e do PE (séculos XIX and XX). *Labor Histórico*, *1*(2), 148–166.

Oesterreicher, W. (1992). Se im Spanischen. Pseudoreflexivität, Diathese und Protoypikalität von semantischen Rollen. *Romanistisches Jahrbuch*, *43*, 287–318.
Ordóñez, F. (1997). *Word order and clause structure in Spanish and other Romance languages*. CUNY.
Ordóñez, Francisco, & Treviño, E. (1999). Left dislocated subjects and the pro-drop parameter: A case study of Spanish. *Lingua*, *107*(1–2), 39–68.
Orfitelli, R., & Hyams, N. (2012). Children's Grammar of Null Subjects: Evidence from Comprehension. *Linguistic Inquiry*, *43*(4), 563–590.
Papafragou, A. (1996). On generics. *University College London Working Papers in Linguistics*, *8*(1), 165–198.
Pierce, A. E. (1992). *Language Acquisition and Syntactic Theory: A Comparative Analysis of French and English Child Grammars*. Springer Netherlands.
Pilati, E., Naves, R., & Salles, H. M. L. (2017). On The Syntax of Subjects in Brazilian Portuguese: Using the 'SPLIT' Pronominal Sytem as the Basis for an Alternative Analysis. *Revista Diadorim*, *19*, 99–139.
Phimsawat, O. (2011). The syntax of pro-drop in Thai. [PhD dissertation]. University of Newcastle.
Rasetti, L. (2000). Null subjects and root infinitives in the child grammar of French. In L. Rizzi & M.-A. Friedemann (Eds.), *Acquistion of Syntax: Studies in comparative developmental linguistics*. Routledge.
Rivero, M. L. (2001). On Impersonal Reflexives in Romance and Slavic and Semantic Variation. In J. Camps & C. R. Wiltshire (Eds.), *Current Issues in Linguistic Theory* (Vol. 216, p. 169). John Benjamins Publishing Company.
Rizzi, L. (1982). *Issues in Italian Syntax:* DE GRUYTER.
Rizzi, L. (1986). Null objects in Italian and the theory of pro. *Linguistic Inquiry*, *17*, 501–557.
Rizzi, L. (1994). Early Null Subjects and Root Null Subjects. In T. Hoekstra & B. D. Schwartz (Eds.), *Language Acquisition and Language Disorders* (Vol. 8, p. 151). John Benjamins Publishing Company.
Rizzi, L. (2005). Grammatically-based target inconsistencies in child language. *The Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition - North America*.
Rizzi, L., & Shlonsky, U. (2007). Strategies of subject extraction. In U. Sauerland & H.-M. Gärtner (Eds.), *Interfaces + recursion = language? Chomsky's minimalism and the view from syntax-semantics*. Mouton de Gruyter.
Roberts, I. (2010). A deletion analysis of null subjects. In T. Biberauer, A. Holmberg, I. Roberts, & M. Sheehan (Eds.), *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge University Press.
Roberts, I., & Holmberg, A. (2010). Introduction: Parameters in minimalist theory. In T. Biberauer, A. Holmberg, I. Roberts, & M. Sheehan (Eds.), *Parametric Variation: Null Subjects in Minimalist Theory*. Cambridge University Press.
Rodrigues, C. (2004). *Impoverish morphology and A-movement out of Case-domain* [Ph.D. dissertation]. University of Maryland.
Roeper, T., & Rohrbacher, B. (2000). Null Subjects in Early Child English and the Theory of Economy of Projection. In S. M. Powers & C. Hamann (Eds.), *The Acquisition of Scrambling and Cliticization* (Vol. 26, pp. 345–396). Springer Netherlands.

Roeper, T., Rooth, M., Mallis, L., & Akiyama, A. (1984). *The problem of empty categories and bound variables in language acquisition*.
Seabrooks, D. (2017). *Child acquisition of Spanish clitic impersonal constructions: An empirical study on the CHILDES corpora* [Master's thesis]. The University of North Carolina at Chapel Hill.
Serratrice, L. (2005). The role of discourse pragmatics in the acquisition of subjects in Italian. *Applied Psycholinguistics*, *26*(3), 437–462.
Sigurðsson, H. Á. (2004). The syntax of Person, Tense, and speech features. *The Italian Journal of Linguistics*, *16*, 219–251.
Sigurðsson, H. Á. (2011). Conditions on Argument Drop. *Linguistic Inquiry*, *42*(2), 267–304.
Sigurðsson, H. Á., & Egerland, V. (2009). Impersonal null-subjects in Icelandic and elsewhere. *Studia Linguistica*, *63*(1), 158–185.
Silva, L. (2017). *As diferenças entre o que se fala e o que se escreve no português do Brasil: A aquisição do clítico se indeterminador e apassivador* [Master's thesis]. Universidade Estadual de Campinas.
Simões, L. J. (1999). Sujeito nulo na aquisição do Português do Brasil. *Cadernos de Estudos Lingüísticos*, *36*, 105–130.
Slobin, D. I. (1973). Cognitive prerequisites for the acquisition of grammar. In C. A. Ferguson (Ed.), *Studies of child language development*.
Slobin, D. I. (1982). Universal and particular in the acquisition of language. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art*. Cambridge University Press.
Tiit-Rein, V. (1997). Fennic. In *The Uralic Languages*. Routledge.
Torn, R. (2002). The status of the passive in English and Estonian. In H. Hendriks (Ed.), *RCEAL Working Papers in Linguistics* (Vol. 7, pp. 81–106).
Torn-Leesik, R. (2007). Voice and modal verbs in Estonian. *Linguistica Uralica*, *43*(3), 173–186. Gale Academic OneFile.
Torn-Leesik, R. (2009). The voice system of Estonian. *STUF - Language Typology and Universals*, *62*(1–2).
Torn-Leesik, R., & Vija, M. (2012). Acquisition of the Impersonal Voice by an Estonian Child. *Journal of Baltic Studies*, *43*(2), 251–271.
Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition*, *40*(1–2), 21–81.
Valian, V. (2016). *Null Subjects* (J. L. Lidz, W. Snyder, & J. Pater, Eds.; Vol. 1). Oxford University Press.
Vija, M., Torn-Leesik, R., & Pajusalu, R. (2009). Tegumood eesti lapsekeeles. *Eesti Rakenduslingvistika Ühingu Aastaraamat. Estonian Papers in Applied Linguistics*, *5*, 329–344.
Wang, Q., Lillo-Martin, D., Best, C. T., & Levitt, A. (1992). Null Subject Versus Null Object: Some Evidence From the Acquisition of Chinese and English. *Language Acquisition*, *2*(3), 221–254.
Weist, R., & Witkowska-Stadink, K. (1985). *Basic relations in child language and word order myth*.
Wexler, K. (1992). Optional infinitives, head movement and the economy of derivation. *Occasional Paper #45, Center for Cognitive Science, MIT*.
Wexler, K. (1995). *Feature-interpretability and optionality in early child grammar*. Workshop on Optionality, Utrecht.

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua*, *106*(1–4), 23–79.
Wexler, K., & Manzini, R. (1987). Parameters and Learnability in Binding Theory. In T. Roeper & E. Williams (Eds.), *Parameter setting*. D. Reidel Pub. Co.; Sold and distributed in the U.S.A. and Canada by Kluwer Academic Publishers.
Wurmbrand, S. (2017). *Icelandic As A Partial Null Subject Language: Evidence From Fake Indexicals*. Zenodo.
Zilles, A. (2002). Grammaticalization of a gente in Brazilian Portuguese. *University of Pennsylvania Working Papers in Linguistics*, *8*(3), Article 22.
Zilles, A. (2007). Grammaticalization of a gente as a cluster of changes: Evidence from apparent and real time studies. *Forúm Linguístico*, *4*(1), 13–4

Appendix A
Sentences and backgrounds used in the survey (Chapter 3)

The order in which the sentences appeared was randomized for each participant who took the survey.

- **ADV (generic)**

Concentração para estudar é um desafio que persegue todos os estudantes. Na USP, há muitas bibliotecas disponíveis e é comum que os alunos estudem lá, procurando por lugares quietos para ler ou fazer um trabalho.
Concentration while studying is a challenge for every student. At USP, there are many libraries available and it is common that students go there to study, searching for a quiet place to read or work on assignments.

Nessa universidade (se) estuda na biblioteca.
In.this university study:3SG in.the library
'At this university one studies in the library.'

- **ADV (existential)**

A mãe do Joaquim tem uma loja chamada 'Só saia'.
Joaquim's mother has a store called 'Only skirts'.

Nessa loja (se) vende saia.
In.this store sell:3SG skirt
'In this store someone sells skirts.'

- **NEG**

Nos Estados Unidos, o bolo de cenoura é feito com uma cobertura de cream cheese, ao invés da cobertura tradicional de chocolate que a gente tem aqui. No Brasil é diferente.
In the US, carrot cakes are made with cream cheese frosting, instead of the traditional chocolate frosting we have here. In Brazil, things are different.

Não (se) faz bolo de cenoura desse jeito.
Not make:3SG cake of carrot in.this way
'One does not make carrot cake in this way.'

Se você era uma criança nos anos 90, você muito provavelmente tinha uma vizinha que vendia geladinho. Mas, as coisas mudaram.
If you were a child in the 90s, you probably had a neighbor who used to sell freeze pops. But then things changed.

Não (se) vende mais geladinho.

Not sell:3SG more freeze.pop
‘One does not sell freeze pops anymore.’

- **NoADV (generic)**

Além do clima frio fazer com que o banho diário não seja uma necessidade para os americanos, alguns acreditam que esse hábito prejudica a pele e os cabelos. No Brasil, é normal e higiênico tomar banhos diariamente.
The cold weather makes daily showers unnecessary for Americans. Besides, some of them believe that daily showers damage the skin and the hair. In Brazil, it is normal and hygienic to shower daily.

(Se) toma banho todo dia.
take:3SG shower every day
‘One showers every day.’

- **NoADV (existential)**

O pai da Mariana tem uma loja chamada ‘Só doce’.
Mariana’s father has a store called ‘Only dessert’.
Mariana’s father has a store called ‘Only dessert’

(Se) vende doce.
sell:3SG dessert
‘Someone sells desserts.’

- **ADV+NEG**

Na Polônia, muitas famílias gostam de colher cogumelos nos bosques, que cobrem mais de 30% do país. Poloneses não acham que é perigoso comer cogumelos selvagens, mas isso não acontece em todas as culturas.
In Poland many families like to pick mushrooms in forests which cover more than 30% of the country. Polish people do not think it is dangerous to eat wild mushrooms, but that does not happen in every culture.

No Brasil não (se) colhe cogumelo em bosques.
In Brazil not pick:3SG mushroom in forests
‘In Brazil one does not pick mushrooms in forests.’

Novelas brasileiras são famosas ao redor do mundo, mas essa popularidade não alcançou todos os países. Em 1995, a guerra da ex-Iugoslávia parou durante 1 semana para que o país assistisse os últimos capítulos da ‘Escrava Isaura’! Mas, como dito, não é em todo lugar que as pessoas assistem novela.

Brazilian soap operas are famous around the world, but this popularity did not reach every country. In 1995, the Yugoslav War stopped for 1 week so that the country could watch the last episodes of 'Slave Isaura'! But, as said before, it is not everywhere that people watch soap operas.

Nos Estados Unidos não (se) assiste novela.
In.the US not watch:3SG soap.operas
'In the US one does not watch soap operas.'

- **ADV+D**

O Joaquim estuda em uma escola em que é obrigatório comer verdura na hora do almoço.
Joaquim studies in a school where it is mandatory to eat vegetables at lunchtime.

Nessa escola tem que (se) comer verdura na hora do almoço.
In.this school have:3SG that eat:INF vegetable at.the time of.the lunch
'In this school one has to eat vegetables at lunchtime.'

Na escola em que a Mariana estuda, tem uma regra em que todo dia os alunos precisam entregar a lição de casa. Quem não entrega a lição de casa, leva suspensão. É uma escola muito rígida.
In the school where Mariana studies, there is a rule which says that students have to turn in homework assignments every day. Who does not do that is suspended from school. This school is very strict.

Nessa escola (se) tem que entregar lição de casa.
In.this school have:3SG that turn.in:INF lesson of home
'In this school one has to turn in homework assignments.'

- **NEG+D**

O abandono escolar é um fenômeno que tira da escola milhares de alunos que podem ser excluídos da sociedade e do mercado de trabalho.
Dropout is a phenomenon that removes several students from school who can be excluded from the society and job market.

Não (se) pode largar a escola.
Not can:3SG leave:INF the school
'One cannot drop out of school.'

De acordo com as leis brasileiras de trânsito, é proibido pedalar na calçada, a não ser que exista sinalização permitindo.
According to the Brazilian traffic regulations, it is forbidden to ride a bicycle on the sidewalk, unless there is a sign allowing it.

Não (se) pode andar de bicicleta na calçada.
Not can:3SG ride:INF of bicycle on.the sidewalk
'One cannot ride a bicycle on the sidewalk.'

- **ADV+NEG+D**

Na escola em que a Mariana estuda, tem uma regra em que os alunos estão proibidos de trazer qualquer animal de estimação.
In the school where Mariana studies there is a rule in which students are forbidden to bring pets.

Nessa escola não (se) pode trazer animal de estimação.
In.this school not can:3SG bring:INF pet
'In this school one cannot bring pets.'

De maridos esquecidos o mundo está cheio, mas Samoa, na Oceania, está tentando melhorar isso. Lá é proibido esquecer o aniversário da esposa, e o marido deve responder à justiça e ainda indenizar a esposa se fizer isso.
The world is full of husbands with poor memory, but Samoa, a country in Oceania, is trying to change it! In Samoa a husband is forbidden to forget his wife's birthday. If he does, he is brought to justice and has to indemnify his wife for his actions.

Em Samoa não (se) pode esquecer o aniversário da esposa.
In Samoa not can:3SG forget:INF the birthday of.the wife
'In Samoa one cannot forget one's wife's birthday.'

- **ANAPHOR (not analyzed)**

No escritório onde o pai do Joaquim trabalha, as pessoas são bastante próximas uma das outras e o clima é muito amigável.
In the office where Joaquim's father works, people are very close to each other and the atmosphere is friendly.

Nesse escritório (se) pode falar de si mesmo.
In.this office can:3SG talk:INF of REFL SELF
'In this office one can talk about oneself.'

- **SUBJECT-ORIENTED ADVERBIALS (not analyzed)**

Escrever tese em casa não é sempre uma boa ideia: o barulho dos vizinhos, de construções acontecendo na rua e tudo o mais podem interferir na concentração. Às vezes, o melhor lugar para escrever uma tese é na biblioteca.

Writing a dissertation at home is not always a good idea: loud neighbors, noise from constructions and everything else can make people distracted. Sometimes the best place to write a dissertation is in the library.

Na biblioteca (se) pode escrever tese em paz.
In.the library can:3SG write dissertation in peace
'In the library one can write a dissertation peacefully.'

- **FILLERS**

O Joaquim está muito cansado depois de um dia na escola em que ele não estudou nada, mas conversou o tempo inteiro com os seus colegas. Quando ele chegou em casa, deitou em sua cama e dormiu tranquilamente por 12 horas.
Joaquim is very tired after a day at school in which he did not study at all, but talked the whole time with his classmates. When he got home, he lay on the bed and slept peacefully for 12 hours.

*A cama foi deitada pelo Joaquim.
The bed was laid.down by Joaquim
'The bed was laid down by Joaquim.'

A maioria das pessoas trabalha todos os dias de segunda à sexta, 8 horas por dia. Para manter uma vida saudável e ter tempo para o lazer e para a família…
Most people work every day from Monday through Friday, 8 hours a day .In order to lead a healthy lifestyle and have time for leisure and family…

No fim de semana é melhor descansar.
On.the weekend is best rest:INF
'It is best to rest on weekends.'

Dentre as pragas que atacam o coco, destaca-se o ácaro da necrose. Essa é uma praga que anda causando muito prejuízo no Brasil. Não tem problema nenhum consumir a água ou a polpa do coco afetado pela praga, mas a aparência do coco dificulta a sua venda.
Among the pests that infest coconut plantations, the coconut mite is the main one. This is a pest that has been causing a lot of financial loss in Brazil. There is no problem consuming the water or the coconut pulp affected by the pest, but the appearance of the coconut makes it difficult to sell.

No Brasil, o ácaro da necrose dificulta a venda do coco.
In Brazil the mite of.the necrosis make.difficult the sales of.the coconut
'In Brazil the coconut mite makes it difficult to sell coconuts.'

A Mariana precisa doar suas roupas para o brechó. Mas, como as roupas estavam um pouco sujas, a Mariana decidiu colocar as roupas na máquina de lavar. Agora elas estão limpas!

Mariana needs to donate her clothes to the thrift store. But as the clothes were a little bit dirty, she decided to wash them in the washing machine. Now the clothes are clean!

? A roupa já lavou pra ser doada no brechó.
The clothes already washed to be donated in.the thrift store
'The clothes were already washed to be donated to the thrift store.'

O Bruno estava brincando na casa do Joaquim com um dinossauro de borracha que tinha dentes muito pontiagudos. Os dois meninos acabaram machucando o dedo.
Bruno was playing at Joaquim's house with a rubber dinosaur that had very sharp teeth. The boys ended up hurting their own fingers.

? O dinossauro do Joaquim$_i$ machucou o Bruno e ele$_i$ com os dentes.
The dinosaur of.the Joaquim hurt the Bruno and him with the teeth
'Joaquim$_i$'s dinosaur hurt Bruno and him$_i$ with his teeth.'

Appendix B
Sentences and slides used in Experiment 1 (Chapter 4)

The following sentences and slides appear in the order they were presented to participants

Slide 1: Você vai assistir alguns vídeos sobre uma escola muito diferente. Lá todas as regras são invertidas.
Now you are going to watch some videos about a very different school. There, all the rules are reversed.

Slide 2: Por exemplo, você deve ter aprendido que a gente tem de escovar os dentes depois de comer, mas nessa escola estranha você não deve escovar os dentes depois de comer.
For example, you must have learned that we have to brush our teeth after eating, but in the weird school you should not brush your teeth after eating.

Slide 3: Tem outras coisas muito estranhas nessa escola: um único aluno tem de comer doce na hora do almoço, enquanto todas as outras crianças comem comida normal.
There are other things in this school that are very weird: only one student has to eat dessert at lunchtime, while all other kids eat regular food.

Slide 4: E eles têm de trazer animais para a escola. Não é estranho?
And they have to bring pets to the school. Isn't weird?

Slide 5: Eu quero que você preste atenção na história e diga se o Elmo está prestando atenção. Ele é um pouco distraído, então às vezes ele não entende direito o que aconteceu na história, mas às vezes ele presta atenção e entende. Como eu nunca sei se ele está prestando atenção ou não, você tem de ajudar ele, tá bom? No final da história, ele vai falar o que ele entendeu e você diz pra ele se isso aconteceu na história ou não.
I want you to pay attention to the story and tell me whether Elmo is paying attention or not. He is a little bit distracted, so sometimes he does not understand what happened in the story, but sometimes he pays attention and understands it. As I never know when he is paying attention or

not, you have to help him, ok? When we finish each story, he is going to tell us what he understood and you should tell him whether that happened in the story or not.

- **Training items**

Slide 6: Essa escola maluca tinha outra regra muito esquisita: ninguém podia dizer a palavra 'dinossauro'.
This school had another very weird rule: no one could say the word 'dinosaur'.

Slide 7: Era dia de trazer brinquedos para a escola e Joaquim trouxe um dinossauro de borracha. A Mariana trouxe uma boneca, o Bruno trouxe uma coleção inteira de carrinhos, e o Lucas trouxe um urso de pelúcia.
It was play day and students had to bring toys to the school. Joaquim brought a rubber dinosaur. Mariana brought a doll. Bruno brought his toy car collection. Lucas brought a teddy bear.

Slide 8: A professora perguntou aos alunos quais brinquedos eles tinham trazido. A Mariana, o Bruno e o Lucas responderam. O Joaquim quase disse a palavra ‘dinossauro’, mas aí ele lembrou da regra esquisita que essa palavra não podia ser dita.
The teacher asked the students which toys they had brought. Mariana, Bruno and Lucas answered the question. Joaquim almost said the word ‘dinosaur’, but then he remembered the weird rule which says that this word should not be pronounced.

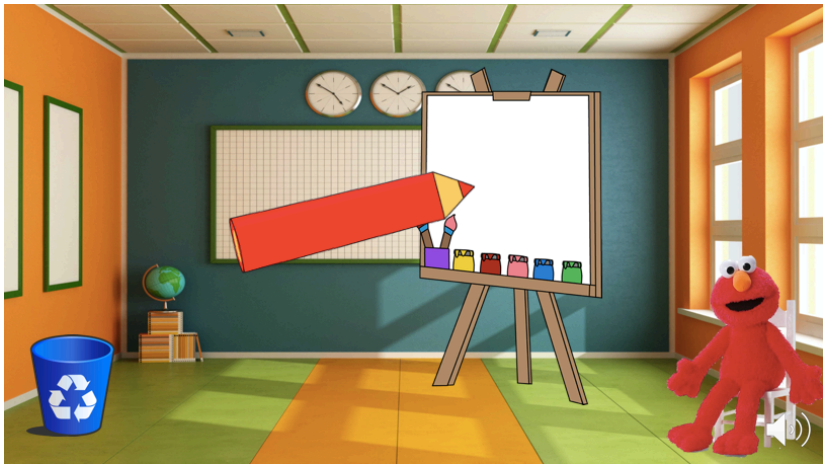
Slide 9: Então, ao invés de falar para a professora que ele trouxe um dinossauro, ele mostrou o brinquedo, sem dizer a palavra.
Therefore, instead of telling the teacher he brought a dinosaur, he showed it to her, without saying the word.

Slide 10: Agora o Elmo vai dizer uma parte da história. Vamos ver se ele prestou atenção ou não.
Now Elmo is going to tell us a part of the story. Let's see whether he paid attention or not.

Slide 11: O Joaquim não disse a palavra dinossauro. (T)rue
The Joaquim not said:3SG the word dinosaur
'Joaquim didn't say the word dinosaur.'

Slide 12: Nessa escola, outra regra esquisita era que os estudantes tinham que usar sempre um lápis de cor vermelha quando eles estavam pintando.

Another weird rule in this school was that students had to always use a red pencil when they were coloring a drawing.



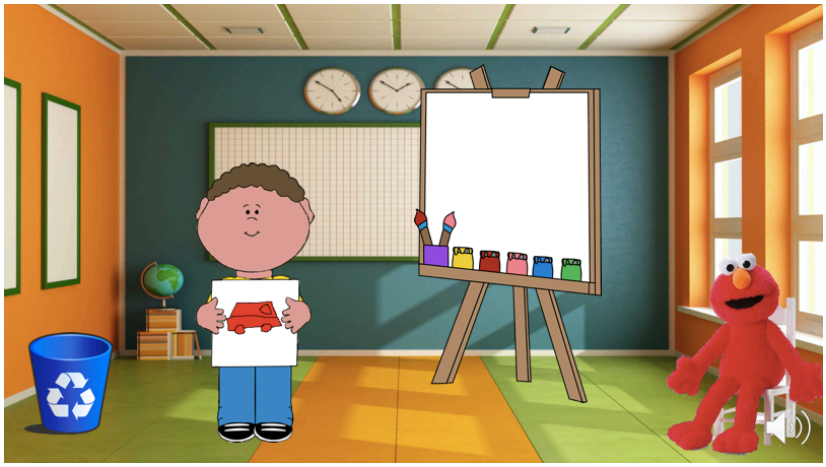
Slide 13: Se eles não usassem a cor vermelha, a professora era obrigada a jogar o desenho no lixo.
If they didn't color a drawing with the red pencil, the teacher had to throw the drawing in the trash can.

Slide 14: A Mariana desenhou uma casa e coloriu ela toda de vermelho.
Mariana drew a house and colored the entire house red.

Slide 15: O Bruno desenhou um gato e coloriu o nariz dele de vermelho.
Bruno drew a cat and colored its nose red.

Slide 16: O Joaquim quase esqueceu dessa regra enquanto estava pintando e quase coloriu o carrinho que ele desenhou todo de amarelo.
Joaquim almost forgot about this rule when he was coloring his drawing. He almost colored the entire car yellow.

Slide 17: Mas, ainda bem que ele lembrou da regra e pintou o carro de vermelho.
But luckily he remembered the rule and colored the car red.



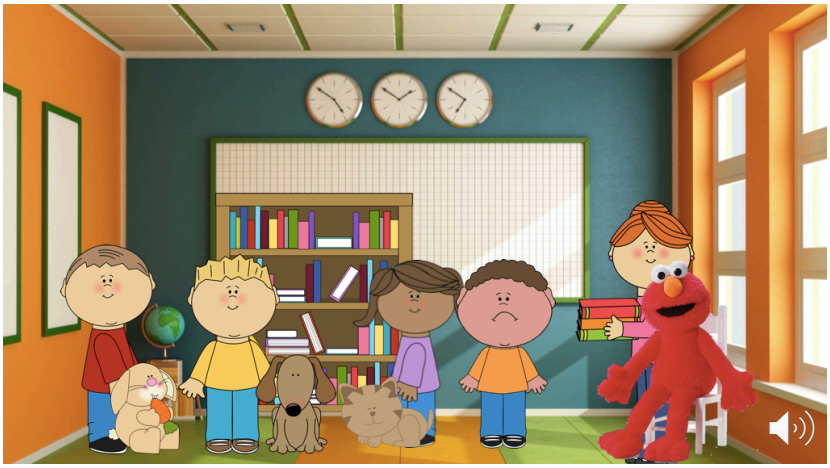
Slide 18: Agora o Elmo vai dizer uma parte da história. Vamos ver se ele prestou atenção ou não.
Now Elmo is going to tell us a part of the story. Let's see whether he paid attention or not.

Slide 19: O Joaquim coloriu o carrinho de amarelo. (F)alse
The Joaquim colored:3SG the little.car of yellow
'Joaquim colored the little car yellow.'

- **Test sentences**

Slide 20: É o primeiro dia de aula. A professora diz: amanhã todo mundo tem que trazer um animal de estimação para a escola, que vai ficar com a gente o ano todo.
It's the first day of school. The teacher says: tomorrow everyone has to bring a pet to school. The pets are going to stay with us for an entire year.

Slide 20: Chegou o dia de trazer animais de estimação pra escola. Alguns alunos quase esqueceram de trazer os animais, mas no fim eles trouxeram. O Bruno trouxe um coelho, a Mariana trouxe um gato, o Lucas trouxe um cachorro, mas o Joaquim esqueceu de trazer um animal de estimação! Coitado do Joaquim!
And came the day when the students had to bring their pets to school. Some students almost forgot to bring their pets, but they ended up remembering to bring them. Bruno brought a bunny, Mariana brought a cat, Lucas brought a dog, but Joaquim forgot to bring a pet! Poor Joaquim!



Slide 21: Agora o Elmo vai dizer uma parte da história. Vamos ver se ele prestou atenção ou não.
Now Elmo is going to tell us a part of the story. Let's see whether he paid attention or not.

Slide 22: Nessa escola *e* tem que trazer animal de estimação. (T)
In.this school have:3SG that bring:INF pets
'In this school one has to bring pets.'

Slide 22 Como você já percebeu, essa escola é muito estranha. Uma das regras é que só um dos alunos, o Joaquim, tem de comer doce na hora do almoço ao invés de comida. Todos os outros alunos comem comida normal.
As you already noticed, this school is very weird. One of the rules is that just one of the students, Joaquim, has to eat dessert during lunch time instead of regular food.

Slide 23 Olha a lancheira do Bruno: ele trouxe uma salada pra comer na hora do almoço. Olha, ele está comendo a salada.
Look at Bruno. He brought a salad to eat for lunch. Look, he is eating the salad now.

Slide 24 A Mariana trouxe pão de queijo pra comer na hora do almoço. Olha, ela está comendo o pão de queijo.
Mariana brought cheese balls to eat for lunch. Look, she is eating cheese balls.

Slide 25 Olha o Lucas, ele está comendo macarrão.
Look at Lucas! He is eating pasta.

Slide 26 O Joaquim primeiro colocou uma banana nanica na lancheira dele, mas lembrou que a regra era que ele tinha de trazer um doce, então ele trouxe um brigadeiro.

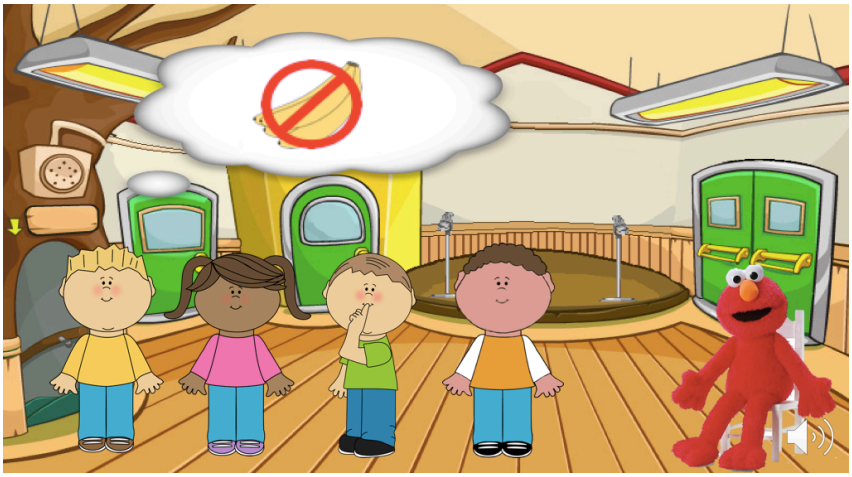
At first, Joaquim put a sandwich in his lunchbox, but then he remembered the rule that he had to bring dessert for lunch, so he brought a brigadeiro instead. Look, he is eating a brigadeiro.

Slide 27 Agora o Elmo vai dizer uma parte da história. Vamos ver se ele prestou atenção ou não.
Now, Elmo is going to tell us a part of the story. Let's see whether he paid attention or not.

Slide 28 Nessa escola *e* tem que comer doce na hora do almoço. (F)
In.this school have:3SG that eat.INF dessert at time of.the lunch
'In this school one has to eat dessert for lunch.'

- **Filler**

Slide 29 Essa escola é tão estranha! Uma das regras absurdas era que os alunos não podiam trazer bananas para a escola na hora do lanche.
This school is so weird! One of the non-sense rules said that students should not bring bananas to school for lunch.



Slide 30 Mas o Joaquim um dia esqueceu dessa regra e colocou uma banana nanica na lancheira dele. Mas aí a mãe lembrou dessa regra e tirou a banana nanica da lancheira do Joaquim, e pôs um bolo no lugar.
But one day Joaquim forgot this rule and put a yellow Cavendish banana in his lunchbox. But then his mother remembered the rule and took the banana out of Joaquim's lunchbox. She put a cake in his lunchbox instead.

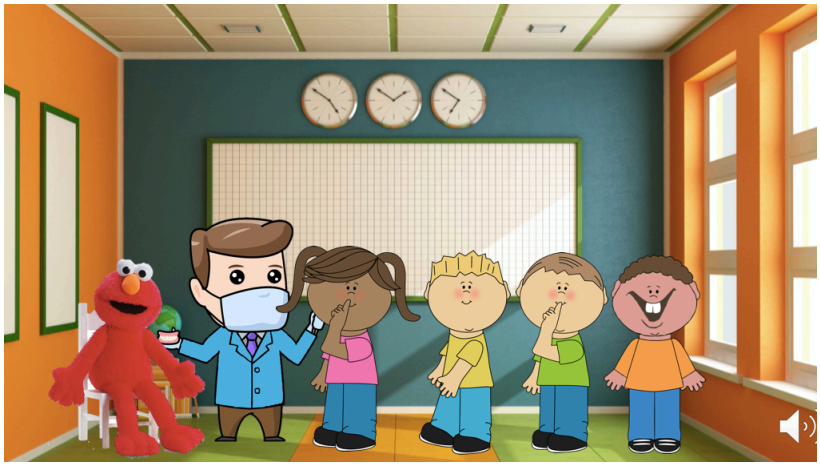
Slide 31 E então ele foi pra escola com um bolo na lancheira.
Then he went to school with a cake in his lunchbox.

Slide 32 Agora o Elmo vai dizer uma parte da história. Vamos ver se ele prestou atenção ou não.
Now, Elmo is going to tell us a part of the story. Let's see whether he paid attention or not.

Slide 33 O Joaquim levou uma banana nanica pra escola. (F)
The Joaquim took:3SG a banana very.small to school
'Joaquim took a yellow Cavendish banana to school.'

- **Test sentences**

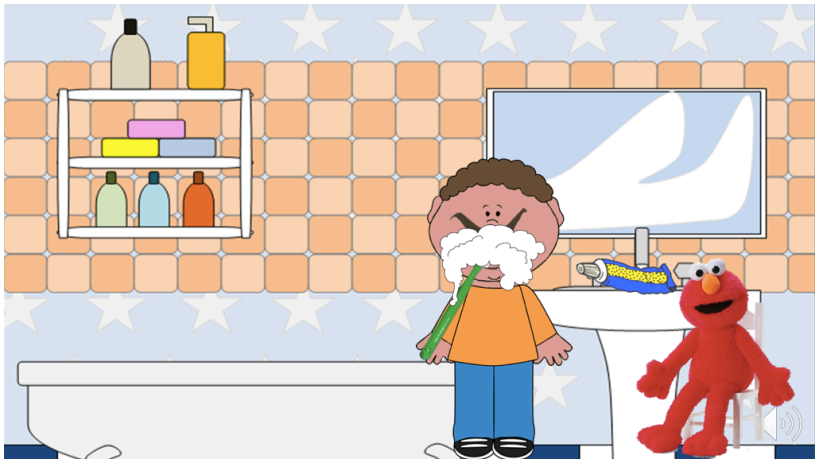
Slide 34 Outra coisa estranha aconteceu nessa escola um dia. Os alunos receberam a visita de um dentista, que conversou com eles. O dentista disse pra todos os alunos que eles não podem escovar os dentes depois de comer, menos o Joaquim, que tinha que escovar os dentes porque eles eram muito grandes.
Another weird thing happened in the school. The students were visited by a dentist. The dentist told the students that they should not brush their teeth after eating, except for Joaquim, who had to brush his teeth because they were too big.



Slide 35 Depois do almoço, os alunos quase esqueceram que eles não podiam escovar os dentes, mas no fim eles lembraram da regra do dentista.
After lunch, the students almost forgot that they should not brush their teeth, but at the end they remembered the rule.

Slide 36 O Joaquim tinha de escovar os dentes dele, porque eram muito grandes. Ele quase esqueceu da recomendação do dentista.
Joaquim had to brush his teeth, because they were too big. He almost forgot about the recommendation given by the dentist.

Slide 37 Mas ele acabou lembrando e foi pro banheiro escovar os dentes depois do almoço.
But he ended up remembering it and went to the restroom to brush his teeth after lunch.



Slide 38 O Bruno foi brincar no pátio depois do almoço sem escovar os dentes. O Lucas também não escovou os dentes depois do almoço e foi brincar junto com o Bruno.
Bruno went to the play yard after lunch without brushing his teeth. Lucas also didn't brush his teeth after lunch and went out to play with Bruno.

Slide 39 A Mariana foi ler um livro e também não escovou os dentes.
Mariana was reading a book and she also didn't brush her teeth.

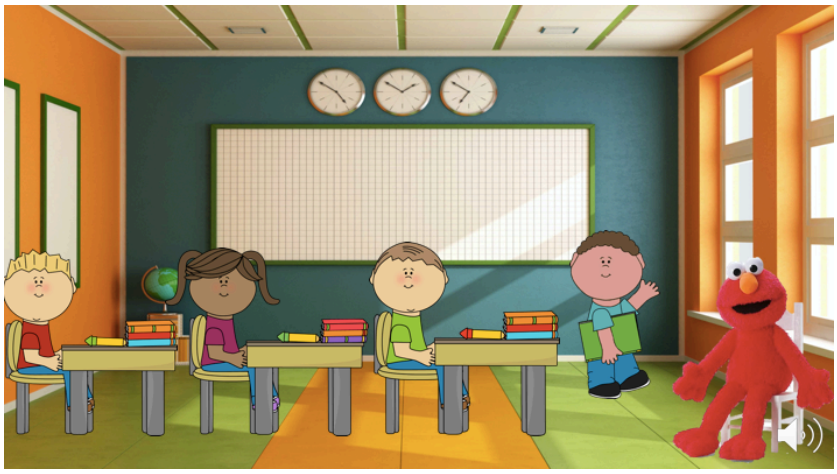
Slide 40 Viu como os estudantes seguiram a regra?
Do you see how the students followed the rule?

Slide 41 Agora o Elmo vai dizer uma parte da história. Vamos ver se ele prestou atenção ou não.
Now Elmo is going to tell us a part of the story. Let's see whether he paid attention or not.

Slide 42: Nessa escola não *e* pode escovar os dentes depois de comer. (T)
In.this school not can:3SG brush:INF the teeth after of eat:INF
'In this school one cannot brush their teeth after eating.'

Slide 43: Essa escola é mesmo muito esquisita. Uma das regras é que enquanto todos os alunos estão estudando de manhã, o Joaquim não pode estudar nessa hora: ele tem de ir para a cama e descansar.
This school is really weird. One of the rules says that while all students are studying in the morning, Joaquim cannot study: he has to go to bed and rest.

Slide 44: Um dia o Joaquim quase esqueceu dessa regra estranha de que ele não pode estudar de manhã e estava preste a abrir um livro, mas ele lembrou da regra.
Once Joaquim almost forgot this weird rule saying that he cannot study in the morning. He was about to open a book, but then he remembered the rule.



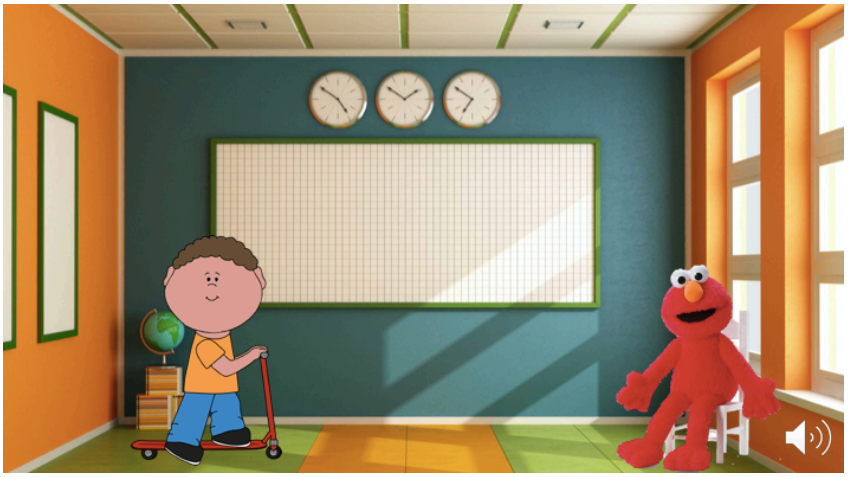
Slide 45: Aí o Joaquim foi descansar, enquanto todos os alunos, a Mariana, o Bruno e o Lucas, estavam estudando.
Then Joaquim rested while all the students - Mariana, Bruno and Lucas - were studying.

Slide 46: Agora o Elmo vai dizer uma parte da história. Vamos ver se ele prestou atenção ou não.
Now Elmo is going to tell us a part of the story. Let's see whether he paid attention or not.

Slide 47: Nessa escola *e* não pode estudar de manhã. (F)
In.this school not can:3SG study:INF of morning
'In this school one cannot study in the morning.'

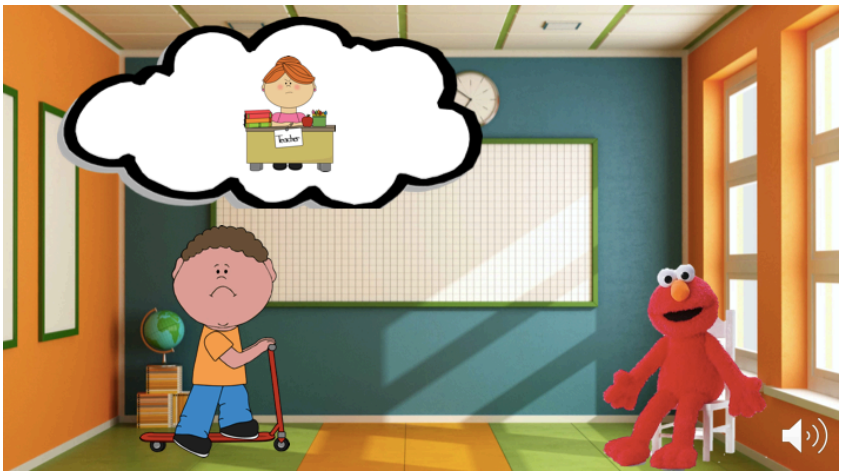
Slide 48: Outra regra muito estranha nessa escola é que todas as crianças tem que brincar de subir nas árvores.
Another very weird rule in this school said that all kids had to climb trees.

Slide 49: Menos o Joaquim que tinha que tinha de brincar dentro da sala de aula.
Except for Joaquim that had to play in the classroom.



Slide 50: O Joaquim está morrendo de vontade de brincar com os seus amigos. Ele queria subir nas árvores com seus amigos.
Joaquim wishes he could play with his friends. He wants to climb the trees with his friends.

Slide 51: Mas ele desistiu da ideia, pensando que a professora ia ficar muito brava com isso.
But he gave up the idea when he thought that the teacher would get very angry about that.



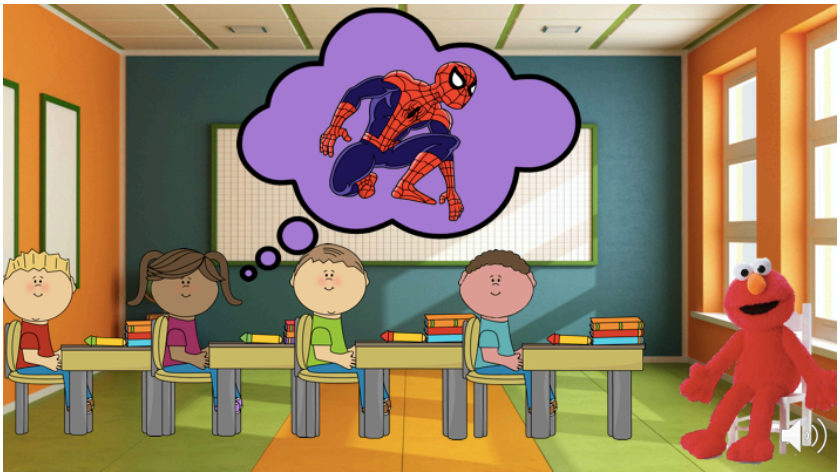
Slide 52: Olha, o Joaquim está brincando dentro da sala de aula!
Look! Joaquim is play in the classroom!

Slide 53 Agora o Elmo vai dizer uma parte da história. Vamos ver se ele prestou atenção ou não.
Now Elmo is going to tell us a part of the story. Let's see whether he paid attention or not.

Slide 54: Nessa escola *e* tem que brincar dentro da sala de aula. (F)
In.this school have:3SG that play:INF inside of.the classroom
'In this school one has to play in the classroom.'

- **Filler**

Slide 55: Outra coisa estranha nessa escola é que todo mundo tinha de vir fantasiado de homem aranha na sexta-feira, até mesmo a Mariana.
Another weird thing about this school is that everyone had to put on a Spiderman costume on Fridays, even Mariana.



Slide 56: O Joaquim não conseguia lembrar se ele tinha de ir fantasiado de Batman ou de Homem Aranha.
Joaquim was unable to remember whether he had to put on a Batman or Spiderman costume.

Slide 57: Ele colocou a fantasia do Batman, mas aí de repente ele lembrou que a regra da escola era ir fantasiado de Homem Aranha.
He put on a Batman costume, but suddenly he remembered that the rule in the school said to put on a Spiderman costume.

Slide 58: Então, ele colocou a fantasia de Homem Aranha…
Then, he put on a Spiderman costume…

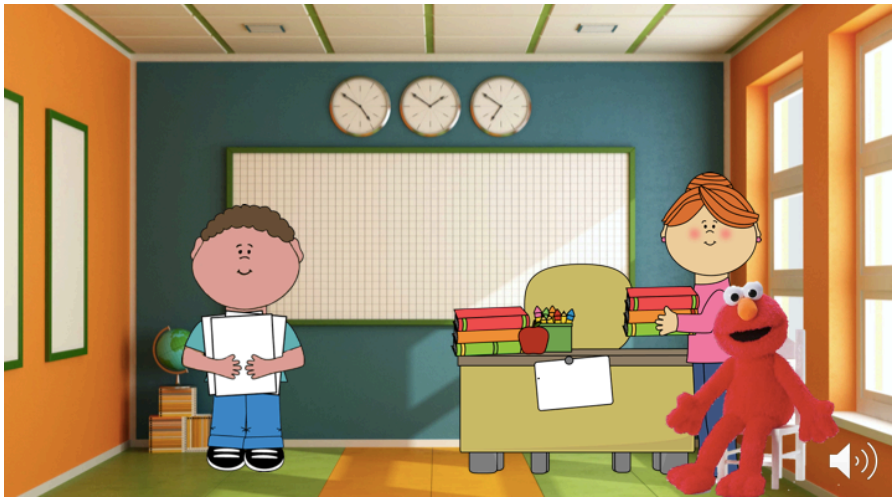
Slide 59: …e foi desse jeito pra escola.
… and he went like that to school.

Slide 60: Agora o Elmo vai dizer uma parte da história. Vamos ver se ele prestou atenção ou não.
Now Elmo is going to tell us a part of the story. Let's see whether he paid attention or not.

Slide 61: O Joaquim foi de Homem Aranha pra escola. (T)
Joaquim went:3SG of Spiderman to school
'Joaquim went to school as Spiderman.'

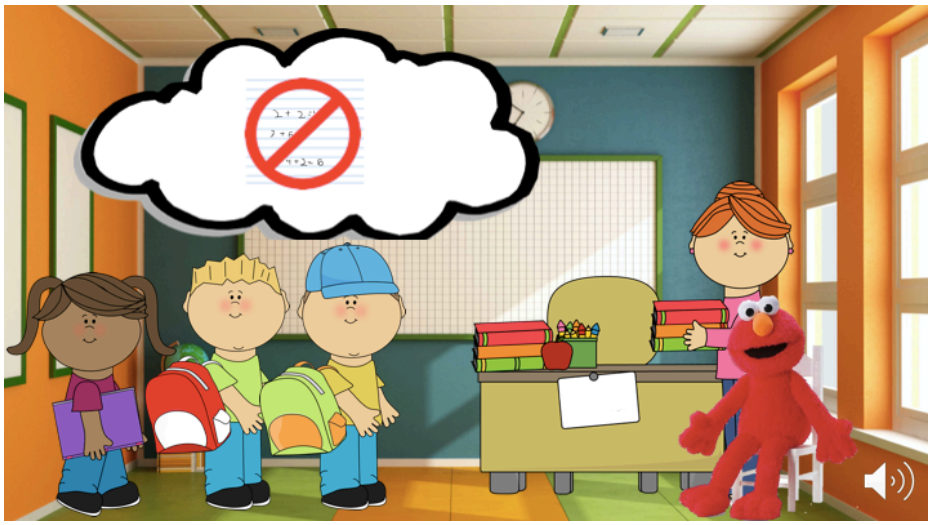
- **Test sentence**

Slide 62: Uma das regras mais estranhas dessa escola é que nenhum aluno deve entregar a lição de casa para a professora, menos o Joaquim, porque ele sempre faz a lição errado.
One of the weirdest rule in this school said that no student should turn in their homework assignment, except for Joaquim because he always does his assignments wrong.



Slide 63: O Joaquim entregou a lição de casa para a professora assim que ele chegou na escola.
Joaquim handed in his homework to the teacher as soon as he got to school.

Slide 64: A Mariana, o Bruno e o Lucas todos fizeram a lição de casa e quase mostraram para a professora, mas no fim eles lembraram dessa regra esquisita e não entregaram a lição de casa.
Mariana, Bruno and Lucas did the homework and they almost showed it to the teacher, but they remembered the weird rule and they didn't turn in their homework assignment.



Slide 65 Agora o Elmo vai dizer uma parte da história. Vamos ver se ele prestou atenção ou não.
Now Elmo is going to tell us a part of the story. Let's see whether he paid attention or not.

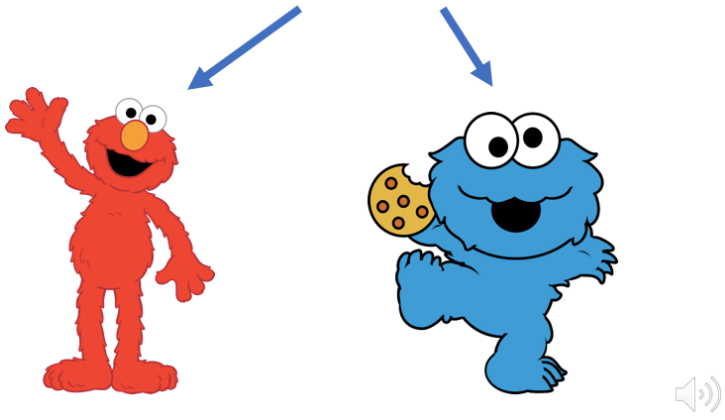
Slide 66: Nessa escola *e* não pode entregar a lição de casa. (F)
In.this school not can:3SG turn.in:INF the lesson of home
'In this school one cannot turn in homework assignments.'

Appendix C

Sentences and slides used in Experiment 2 (Chapter 5)

The following sentences and slides appear in the order they were presented to participants. When Elmo talks, the arrow points to him, when Cookie Monster talks, the arrow points to him instead.

Slide 1: A gente vai assistir umas histórias com dois personagens: o Elmo e o Cookie Monster. No final de cada história, eles vão dizer uma parte da história. Nessa brincadeira, você tem de me ajudar a descobrir qual personagem disse do melhor jeito uma parte da história. Porque às vezes o Elmo ou o Cookie Monster dizem coisas que não fazem muito sentido ou parecem esquisitas. Às vezes também acontece que os dois fantoches dizem coisas do jeito certo e que fazem sentido. Também acontece que às vezes os dois dizem algo errado ou esquisito. Você tem de prestar atenção nisso, ok? Vamos começar a brincadeira! Preste atenção nos vídeos e no que cada fantoche vai dizer.
We are going to watch some stories with two characters: Elmo and Cookie Monster. At the end of each story, they are going to tell us a part of the story. In this game, you have to help me to find out which character said in the best way a part of the story. I'm asking you to do that because sometimes Elmo or Cookie Monster say things that do not make sense or that seem weird. Sometimes it also happens that both of them say things in the right way and that make sense. It also happens that sometimes none of them say something wrong or weird. You have to pay attention to this, ok? Let's start the game! Pay attention to the videos and to what each character is going to say.



- **Training items**

Slide 2: Os alunos tiveram lição de casa e hoje a professora está recolhendo as lições.
The students had a homework assignment and the teacher is collecting them.

Slide 3: A Marina está procurando na mochila dela a lição que ela fez. Será que ela consegue encontrar?
Mariana is looking in her backpack for the assignment she did. Do you think she is going to find it?

Slide 4: Olha só! Ela encontrou! E agora ela está entregando a lição pra professora.

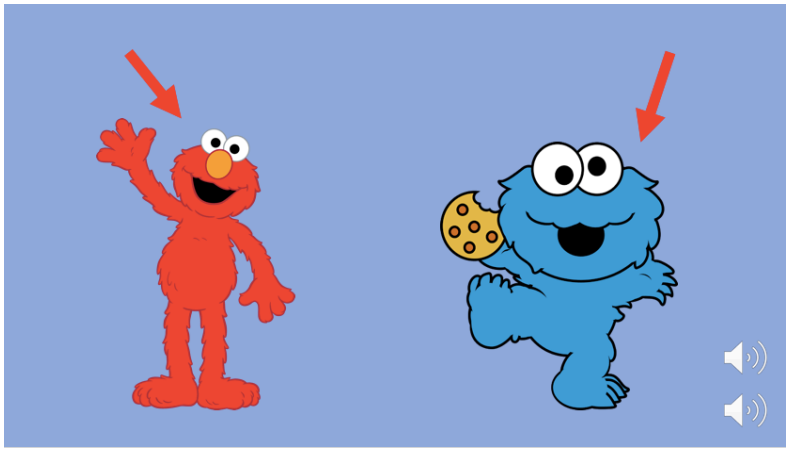
Oh, look! She found it! Now she is turning in her homework.

Slide 5: Agora vamos ver se tem um personagem que diz do melhor jeito alguma coisa que aconteceu na história.
Now let's see if there is a character that says in the best way something that happened in the story.

Slide 6:

Elmo:	A	Mariana	fez	a	lição	de casa. (Both)
	The	Mariana	did:3SG	the	lesson	of home
	'Mariana did the homework.'					

CM:	A	Mariana	entregou	a	lição	de casa.
	The	Mariana	turned.in:3SG	the	lesson	of home
	'Mariana turned in the homework.'					



Slide 7: É dia de trazer animais de estimação pra escola. O Joaquim estava pensando em trazer um coelho.
It is day to bring pets to school. Joaquim was thinking about bringing a bunny.

Slide 8: Mas no final ale decidiu trazer uma tartaruga pra escola.
But he ended up deciding to bring a turtle to school.

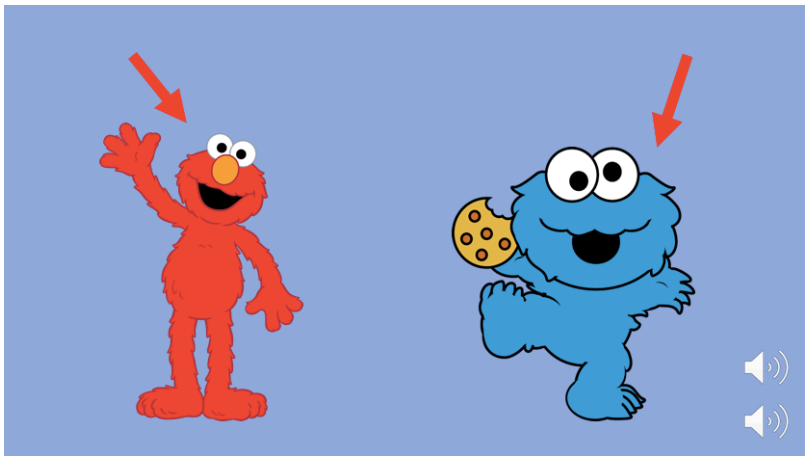
Slide 9: Olha! Todo mundo está brincando com a tartaruga do Joaquim!
Look! Everyone is playing with Joaquim's turtle!

Slide 10: Agora vamos ver se tem um personagem que diz do melhor jeito alguma coisa que aconteceu na história.
Now let's see if there is a character that says in the best way something that happened in the story.

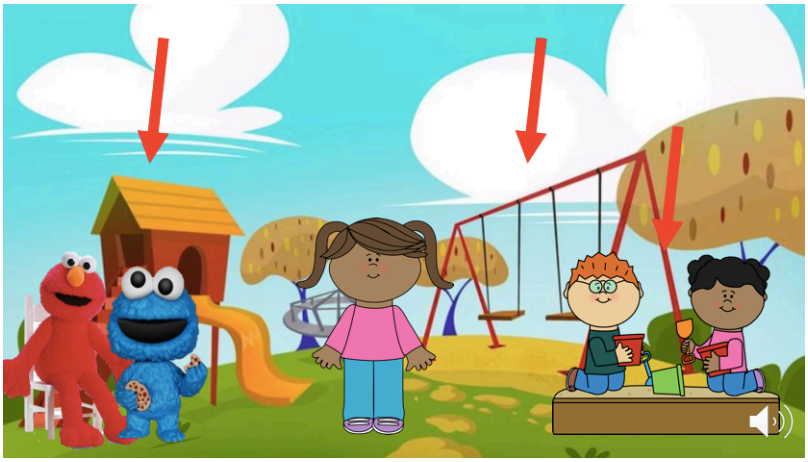
Slide 11:

Elmo: O Joaquim trouxe uma tartaruga pra escola. (Elmo)
The Joaquim brought:3SG a turtle to school
'Joaquim brought a turtle to school.'

CM: O Joaquim pode ter trazido uma tartaruga pra escola.
The Joaquim might:3SG have:INF brought a turtle to school
'Joaquim might have brought a turtle to school.'



Slide 12: Os alunos estão se divertindo no pátio. A Mariana está em dúvida se ela quer brincar na areia, no escorregador ou no balanço.
The students are having fun in the play yard. Mariana is thinking whether she wants to play in the sandbox, on the slide or on the swing.



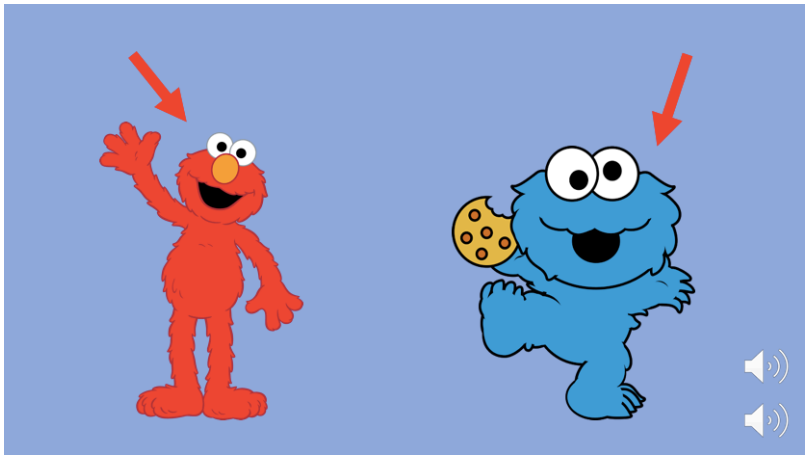
Slide 13: A Mariana chamou o Bruno pra brincar no balanço com ele.
Mariana invited Bruno to play on the swing with him.

Slide 14: Olha lá! Ela está brincando no balanço!
Look! She is playing on the swing!

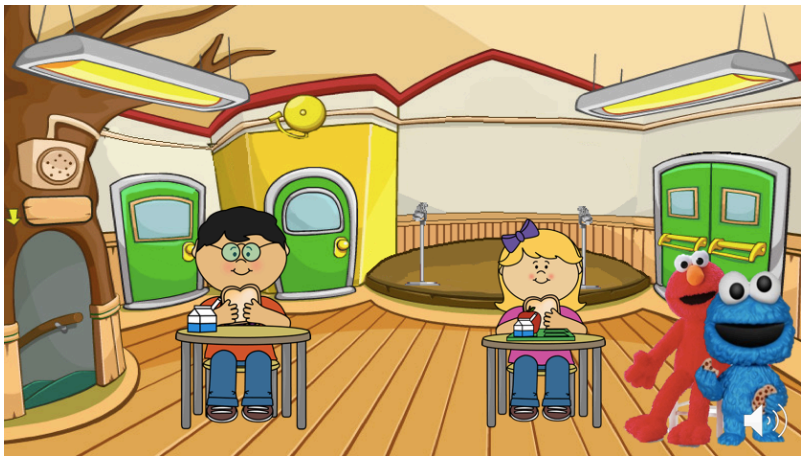
Slide 15: Agora vamos ver se tem um personagem que diz do melhor jeito alguma coisa que aconteceu na história.
Now let's see if there is a character that says in the best way something that happened in the story.

Slide 16: *Elmo:* A Mariana brincou na areia. (Neither)
The Mariana played in.the sand
'Mariana played in the sandbox.'

CM: A Mariana foi no escorregador.
The Mariana went on.the slide
'Mariana played on the slide.'



Slide 17: As crianças agora estão almoçando.
The kids are having lunch now.



Slide 18: Olha só, o Joaquim vai beber leite.
Look! Joaquim is going to drink milk.

Slide 19: Ah não, o Joaquim derramou leite no chão!
Oh no, Joaquim spilled the milk!

Slide 20: Coitado! Agora ele está chorando.
Poor Joaquim! Now he is crying.

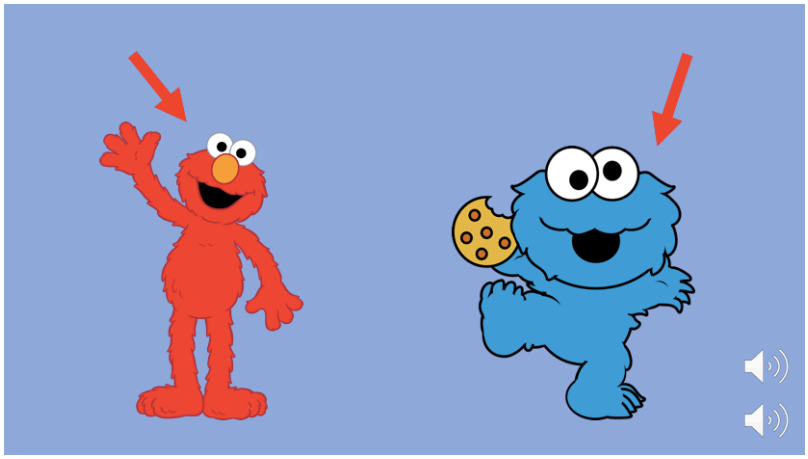
Slide 21: A professora falou pro Joaquim que ele não precisa chorar por causa do leite derramado.
The teacher told Joaquim he does not need to cry over the spilled milk.

Slide 22: Agora vamos ver se tem um personagem que diz do melhor jeito alguma coisa que aconteceu na história.
Now let's see if there is a character that says in the best way something that happened in the story.

Slide 23:

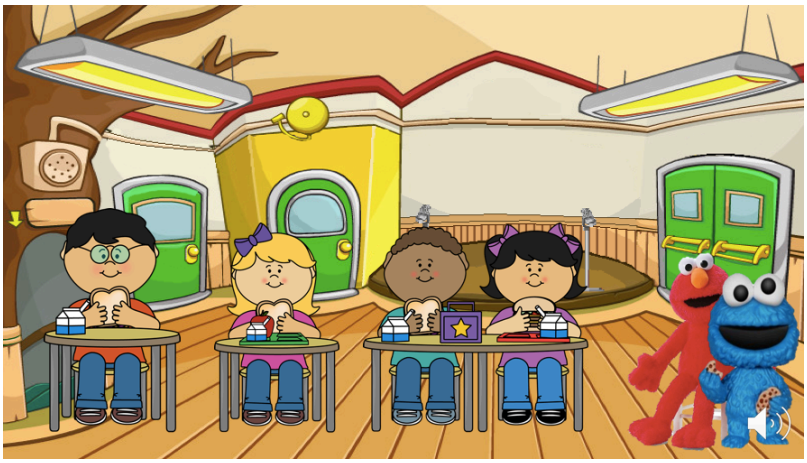
Elmo: O Joaquim comeu um ovo. (CM)
The Joaquim ate:3SG an egg
'Joaquim ate an egg.'

CM: O Joaquim chorou.
The Joaquim cried:3SG
'Joaquim cried.'



- **Test sentence (NoADV)**

Slide 24: Nessa escola, os alunos comem almoço todo dia. Olha eles comendo almoço agora!
In this school, students eat lunch every day. Look at them eating lunch now!



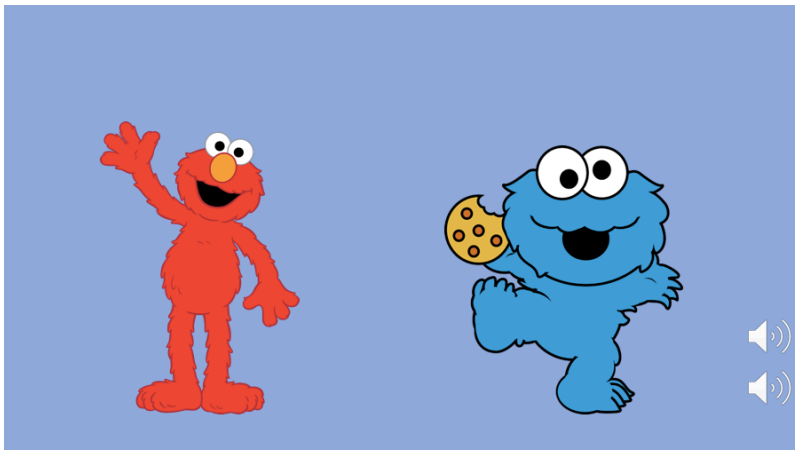
Slide 25: Agora vamos ver se tem um personagem que diz do melhor jeito alguma coisa que aconteceu na história.
Now let's see if there is a character that says in the best way something that happened in the story.

Slide 26:

Elmo: e come almoço. (Neither)
eat:3SG lunch
'One eats lunch.'

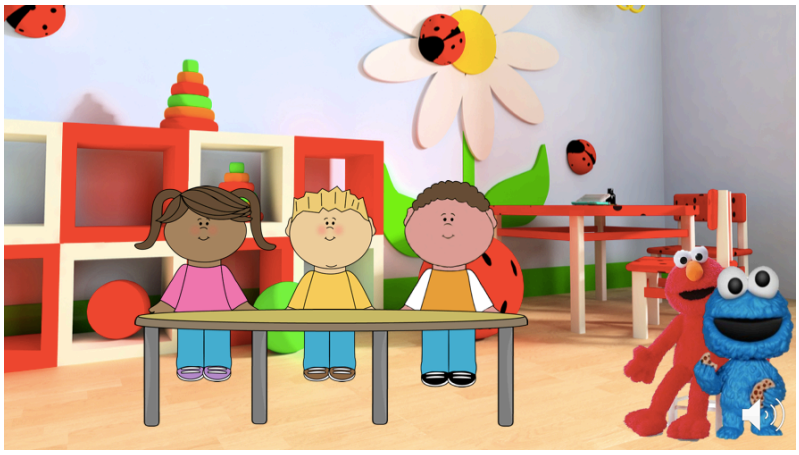
CM: Se come almoço.

SE eat:3SG lunch
‘One eats lunch.’



- **Fillers**

Slide 27: A Mariana, o Bruno e o Joaquim estão jogando um jogo.
Mariana, Bruno and Joaquim are playing a game.



Slide 28: O Joaquim precisa adivinhar em que mão está um anel.

Slide 29: Talvez o anel esteja na mão do Bruno ou da Mariana.
The ring might be in Bruno's or Mariana's hand.

Slide 30: Agora vamos ver se tem um personagem que diz do melhor jeito alguma coisa que aconteceu na história.
Now let's see if there is a character that says in the best way something that happened in the story.

Slide 31:

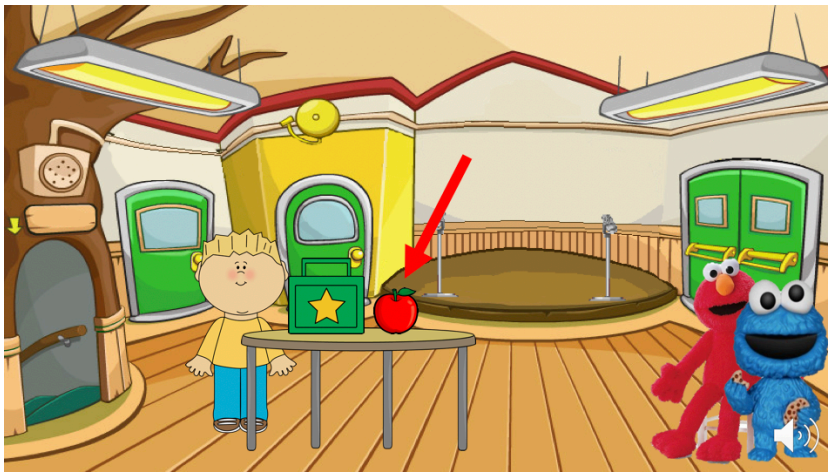
Elmo:	A	Mariana pode	ter	o	anel.	(Elmo)
	The	Mariana might:3SG	have:INF	the	ring	
	'Mariana might have the ring.'					

CM:	A	Mariana tem que	ter	o	anel.	
	The	Mariana have that	have:INF	the	ring	
	'Mariana has to have the ring.'					

Slide 32:

O Bruno trouxe alguma coisa pro lanche. Vamos ver o que está na lancheira dele.
O Bruno brought something for lunch. Let's see what is in his lunchbox.

Slide 33:
Olha! Uma maçã!
Look! An apple!

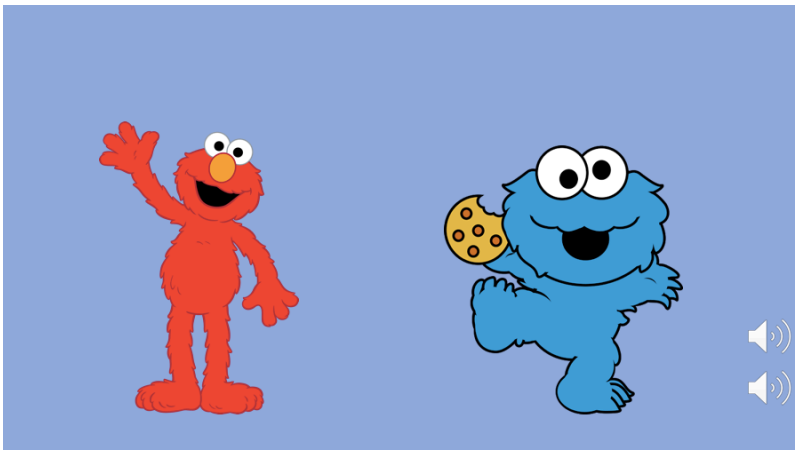


Slide 34: Agora vamos ver se tem um personagem que diz do melhor jeito alguma coisa que aconteceu na história.
Now let's see if there is a character that says in the best way something that happened in the story.

Slide 35:

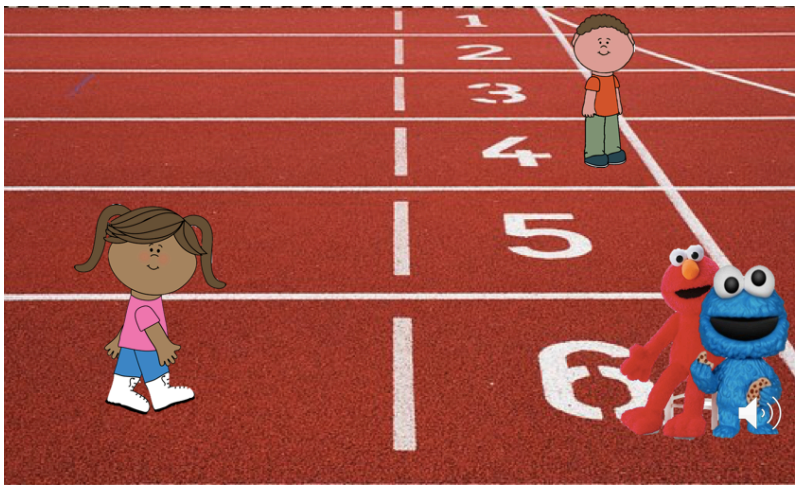
Elmo:	O	Bruno pode	ter	uma maçã.	(CM)
	The	Bruno might:3SG	have:INF	an apple	
	'Bruno might have an apple.'				

CM:	O	Bruno tem	que ter	uma	maçã.
	The	Bruno have:3SG	that have:INF	an	apple
	'Bruno has to have an apple.'				



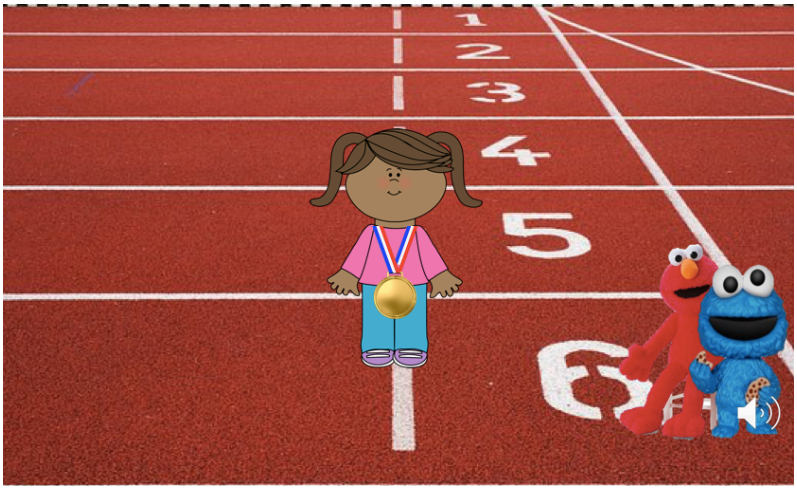
- **Test sentence (ADV+EP)**

Slide 36: One day, the school was having a competition to see who could run faster. Joaquim almost wins the competition…



Slide 37: But, Mariana ended up winning it.

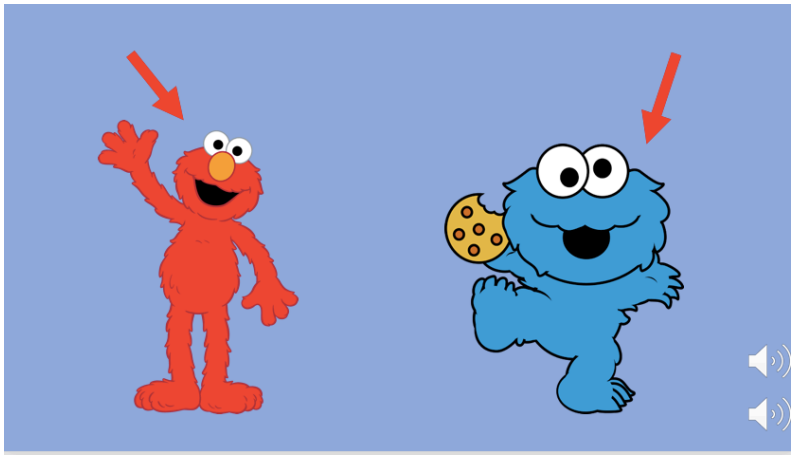
Slide 38: Mariana earned a medal.



Slide 39: Now let's see if there is a character that says in the best way something that happened in this story.

Slide 40: *Elmo:* Nessa escola *e* pode ganhar uma medalha. (CM = se)
In.this school may earn.INF a medal
'In this school one may earn a medal.'

CM: Nessa escola se pode ganhar uma medalha.
In.this school SE may earn.INF a medal
'In this school one can earn a medal.'

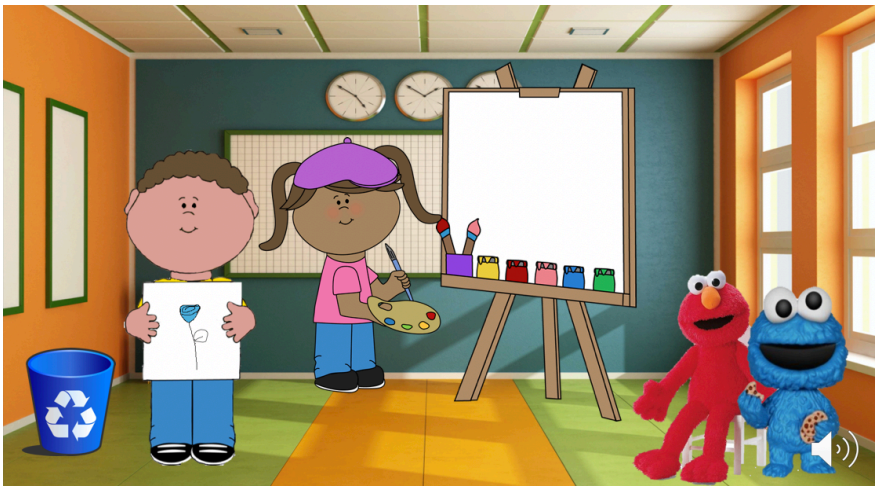


- **Filler**

Slide 41: Os estudantes estão desenhando e pintando. O Joaquim desenhou uma flor muito bonita. Ele adora desenhar. Ele pensou em colorir a flor de Vermelho, mas desistiu da ideia, pensando que uma flor azul seria muito mais original e bonita.
The students are drawing and coloring. Joaquim draw a very beautiful flower. He loves drawing. He thought of coloring the flower red, but he gave up the idea. He thought that a blue flower would be more unique and pretty.

Slide 42: Olha, o Joaquim está pintando a flor de azul.
Look, Joaquim is coloring the flower blue.

Slide 43: Olha como a flor azul ficou bonita!
Look how pretty is the blue flower!

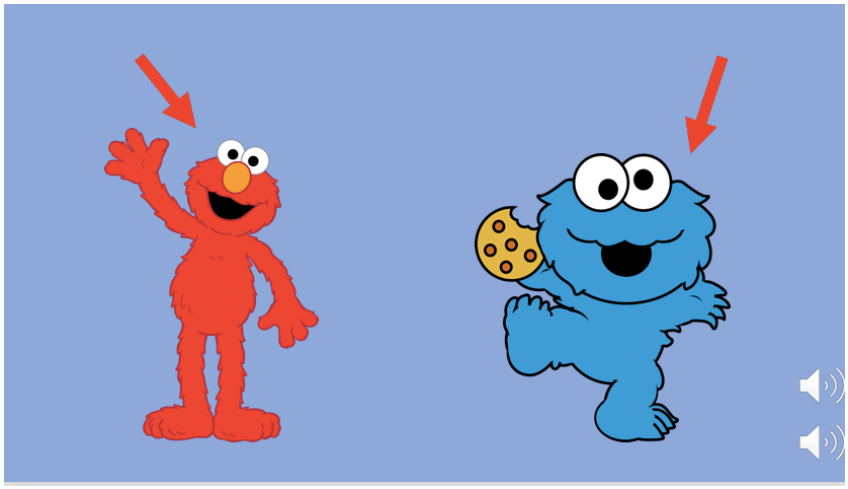


Slide 44: Agora vamos ver se tem um personagem que diz do melhor jeito alguma coisa que aconteceu na história.
Now let's see if there is a character that says in the best way something that happened in the story.

Slide 45:

Elmo: O	Joaquim	pintou	a	flor	de vermelho.	(CM)
The	Joaquim	colored	the	flower	of red	
'Joaquim colored the flower red.'						

CM: O	Joaquim	pintou	a	flor	de azul.
The	Joaquim	colored	the	flower	of blue
'Joaquim colored the flower blue.'					



Test sentence (ADV)

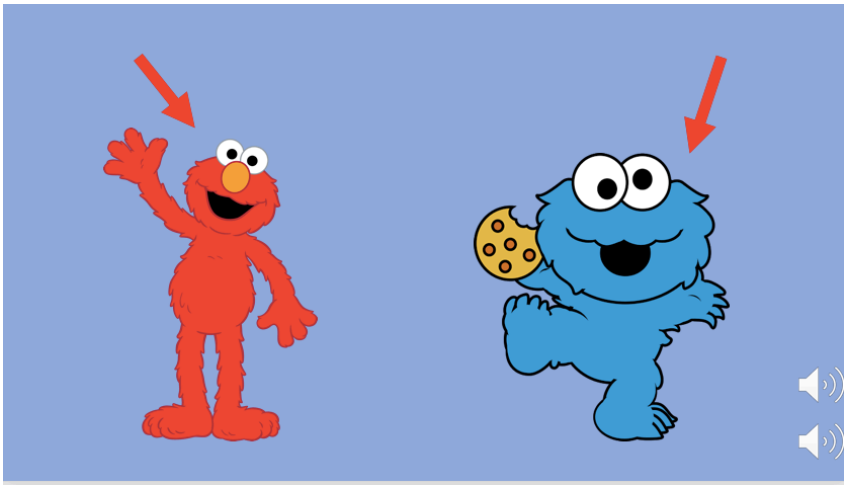
Slide 46: Nessa escola os alunos podem trazer brinquedo uma vez por semana. Hoje é dia de trazer brinquedo. Olha como as crianças estão se divertindo.
In this school students can bring toys once a week. Today is play day. Look how the kids are having fun!

Slide 47: Agora vamos ver se tem um personagem que diz do melhor jeito alguma coisa que aconteceu na história.
Now let's see if there is a character that says in the best way something that happened in the story.

Slide 48:

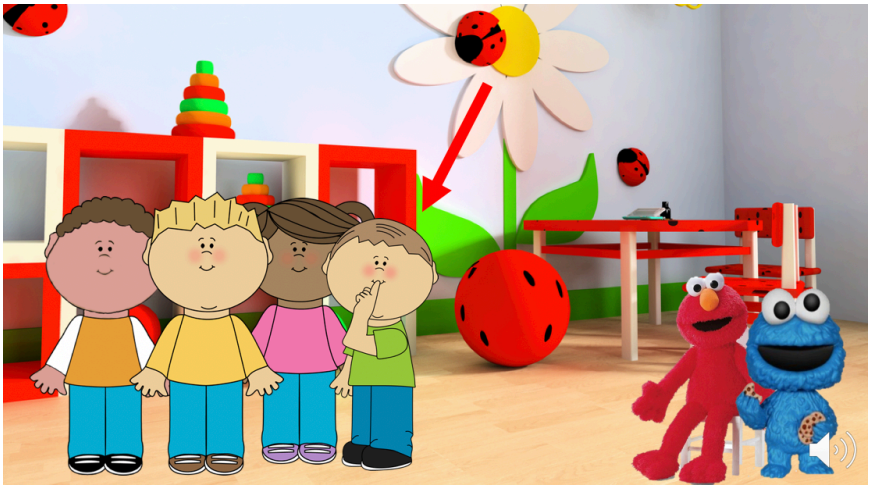
Elmo: Nessa escola se traz brinquedo. (Elmo = se)
In.this school SE bring:INF toy
'In this school one brings toys.'

CM: Nessa escola *e* traz brinquedo.
In.this school bring:INF toy
'In this school one brings toys.'



- **Filler**

Slide 49: É dia de trazer brinquedo pra escola. Vamos ver o que o Lucas trouxe.
It is day to bring toys to school. Let's see what Lucas brought to school.



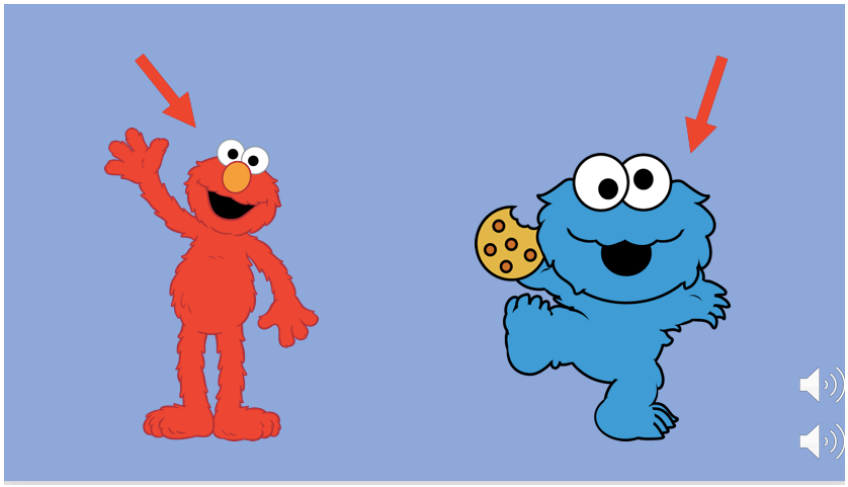
Slide 50: Olha, ele trouxe o porco de pelúcia dele. Você já viu um porco de pelúcia antes?
Look, he brought a stuffed pig. Have you seen a stuffed pig before?

Slide 51: Agora vamos ver se tem um personagem que diz do melhor jeito alguma coisa que aconteceu na história.
Now let's see if there is a character that says in the best way something that happened in the story.

Slide 52:

Elmo: O Lucas pode ter um porco de pelúcia. (CM)
The Lucas might:3SG have:INF a pig of plush
'Lucas might have a stuffed pig.'

CM: O Lucas tem que ter um porco de pelúcia.
The Lucas have:3SG that have:INF a pig of plush
'Lucas has to have a stuffed pig.'



- **Test sentence (ADV+D)**

Slide 53: Uma regra estranha nessa escola é que os estudantes não podem comer banana no almoço.
A weird rule in this school is that the students cannot eat bananas for lunch.

Slide 54: O Bruno quase esqueceu dessa regra e estava prestes a comer uma banana.
Bruno almost forgot this rule and was about to eat a banana.

Slide 55: Mas então ele lembrou e comeu uma laranja ao invés de comer uma banana.
But then he remembered it and ate an orange instead.

Slide 56: *Agora vamos ver se tem um personagem que diz do melhor jeito alguma coisa que aconteceu na história.*
Now let's see if there is a character that says in the best way something that happened in this story.

***Slide* 57:**

Elmo: Nessa escola não *e* pode comer banana. (Both)
In.this school not can eat.INF banana
'In this school one cannot eat bananas.'

CM: Nessa escola não *se* pode comer banana.
In.this school not SE can eat.INF banana
'In this school one cannot eat bananas.'